DIAGNOSTIC ASSESSMENT AND ADVANCEMENT OF MULTI-OBJECTIVE RESERVOIR CONTROL UNDER UNCERTAINTY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jazmin Zatarain Salazar

August 2018

DIAGNOSTIC ASSESSMENT AND ADVANCEMENT OF MULTI-OBJECTIVE RESERVOIR CONTROL UNDER UNCERTAINTY

Jazmin Zatarain Salazar, Ph.D.

Cornell University 2018

This dissertation contributes to the assessment of new scientific developments for multiobjective decision support to improve multi-purpose river basin management. The main insights of this work highlight opportunities to improve modeling of complex multi-purpose water reservoir systems and opportunities to flexibly incorporate emerging demands and hydro-climatic uncertainty. Additionally, algorithm diagnostics contributed in this work enable the water resources field to better capitalize on the rapid growth in computational power. This opens new opportunities to increase the scope of the problems that can be solved and contribute to the robustness and sustainability of water systems management worldwide. This dissertation focuses on a multi-purpose reservoir system that captures the contextual and mathematical difficulties confronted in a broad range of global multi-purpose systems challenged by multiple competing demands and uncertainty.

The first study demonstrates that advances in state of the art multiobjective evolutionary optimization enables to reliably and effectively find control policies that balance conflicting tradeoffs for multi-purpose reservoir control. Multiobjective evolutionary optimization techniques coupled with direct policy search can reliably and flexibly find suitable control policies that adapt to multi-sectorial water needs and to hydro-climatic uncertainty. The second study demonstrates the benefits of cooperative parallel MOEA architectures to reliably

and effectively find many objective control policies when the system is subject to uncertainty and computational constraints. The more advanced cooperative, co-evolutionary parallel search expands the scope of problem difficulty that can be reliably addressed while facilitating the discovery of high quality approximations for optimal river basin tradeoffs. The insights from this chapter should enable water resources analysts to devote computational efforts towards representing reservoir systems more accurately by capturing uncertainty and multiple demands when properly using parallel coordinated search. The third study extended multi- purpose reservoir control to better capture flood protection. A risk-averse formulation contributed to the discovery of control policies that improve operations during hydrologic extremes. Overall this dissertation has carefully evaluated and advanced the Evolutionary Multiobjective Direct Policy Search (EMODPS) framework to support multi-objective and robust management of conflicting demands in complex reservoir systems.

BIOGRAPHICAL SKETCH

Jazmin Zatarain Salazar grew up in Texcoco, México. She attended the University of Chapingo, México were she received a B.S. in Agro-industrial Engineering in 2009. She then received an M.S. in Environmental Science from the University of San Luis Potosí, México in 2010; and a M.S. in Natural Resources Management from Colognne University of Applied Sciences, Germany in 2011. She then began a Ph.D. in Environmental and Water Resources Systems Engineering in the Department of Civil and Environmental Engineering at Cornell University. She will begin a postdoctoral research position in the Department of Electronics, Information and Bioengineering at Politecnico di Milano, Italy in September 2018.

This dissertation is dedicated to my husband Alberto, who is my world and who has held me everyday through graduate school; and to my wonderful parents Alfredo and Raquel who taught me the value of education and have selflessly supported me in every step in life.

ACKNOWLEDGEMENTS

I would first like to thank my advisor Prof. Patrick Reed for patiently holding me through the path of geekdom, also for believing in me and for having the most profound impact in my academic growth. His enthusiasm, generosity and drive produce the best Pareto approximate mentorship that one could wish for. I would also like to thank my committee members Prof. Oliver Gao, Prof. Todd Walter and Prof. Lindsay Anderson for their commitment in making every avenue possible for successfully reaching my milestones.

Additionally, I would like to thank my brilliant collaborators Prof. Jon Herman, Prof. Julianne Quinn, Prof. Matteo Giuliani and Prof. Andrea Castelletti for their encouragement and for generously sharing their knowledge and expertise. Also, to Dave Hadka, for creating the tools that make us all more productive. To my fellow EWRSs and my friends in Ithaca who have become my family from afar.

I also want to thank the people of México who through CONACYT supported my graduate studies. Portions of this work were supported by Consejo Nacional de Ciencia y Tecnologia (CONACYT) Fellowship No. 313591; and by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507.

Finally, I am incredibly lucky to have the support of my amazing family and friends. I am especially thankful for Alberto who dedicates every moment in our marriage making me happy and whose love and support give me the strength to take on any endeavor; and my loving parents who are my source of inspiration and who have worked hard to give me every opportunity to succeed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Addressing Challenges in Multipurpose Reservoir Operations to Improve River Basin Management

Reservoirs play an important role in basin-wide water resources management. Globally, hydropower reservoirs provide services beyond electricity generation; they often need to balance conflicting demands for water supply, flood protection, environmental and recreational services. Carefully crafting these reservoir control policies is of paramount importance since they dictate multi-decadal operations leaving rare opportunity of readdressing operations to consider expanding water demands. Recent advances in *a posteriori* multi-objective decision support can aid in discovering an explicit representation of the system's tradeoffs (Tsoukiàs, 2008). This approach has demonstrated benefits on policy design by moving away from highly constrained and aggregated formulations that concede to the decision-makers preconceptions of the system, strongly limiting the discovery of tradeoffs and operation alternatives. A multi-objective, *a posteriori* approach provides a flexible set of alternatives, (Cohon and Marks, 1973, 1975a; Messac and Mattson, 2002; Reynoso-Meza et al., 2014), by generating the full set of tradeoffs for the system (Cohon et al., 1979). Moreover, the discovery of suitable policy designs that improve performance across multiple objectives involves coupling the analysis with high dimensional visualization techniques (Kollat et al., 2011; Kollat and Reed, 2007; Woodruff et al., 2013), allowing interaction and stakeholder feedback by exploring tradeoffs and their design consequences (Roy, 1971, 1990; Brown et al., 2015). This style of analy-

sis has been widely adopted in water resources literature (Nicklow et al., 2010) and has recently permeated water agencies for practical planning (Basdekas, 2014; Brown et al., 2015). Nevertheless, significant effort is still required to bridge truly multi-purpose reservoir control policy design and its actual implementation for reservoir operations. Sophisticated solution techniques are required to solve a multi-purpose abstraction of the reservoir system that captures their vulnerability across a broad suite of hydro-climatic states. This is still an emerging area of research and widespread acceptance in practice is yet to be solidified. This dissertation utilizes Evolutionary Multiobjective Direct Policy Search (EMODPS) due to its proven capability to solve challenging multiobjective control problems (Giuliani et al., 2014a, 2015b; Quinn et al., 2017a; Koutsoyiannis and Economou, 2003). Succinctly, the EMODPS approach uses global non-linear approximators to parameterize candidate operating policies and multi-objective evolutionary algorithm (MOEAs) to find Pareto approximate policies over the problem's conflicting objectives. This method can easily include multiple state variables and simulate them over a range of stochastic inputs during the optimization without building an explicit transition probability model, removing limits to the number of exogenous variables, such as streamflow and precipitation, that can be used to condition reservoir release decisions, and the number of Pareto-optimal solutions that can be discovered by simultaneously optimizing performance across multiple objectives. EMODPS is a promising alternative for discovering robust multi-purpose reservoir control policies; however, its success is highly dependent upon 1) the capability of non-linear approximators to accurately represent the policy in a multiobjective structure, Giuliani et al. (2015c) has demonstrated that radial basis functions (RBFs) are well suited to approximate multipurpose control policies, 2) finding

the balance between the amount and type of information to shape the control policy; in other words, we want to shape a policy that takes advantage of the systems information while also achieving a generalizable policy that is able to perform well under broader conditions, and 3) the capability of MOEAs to support search of Pareto approximate policies for the system's conflicting tradeoffs. Multi-purpose reservoir control problems result in higher-dimensional, many-objective spaces. It is therefore necessary to employ tools that can effectively search the high-dimensional problem formulations and capture potentially complex tradeoffs between the sectorial demands. The resulting problem formulation features challenging properties including (1) many-objective formulations, (2) multi-modality, (3) nonlinearity, (4) discreteness, (5) severe constraints, (6) stochastic objectives, and (7) non-separability (Reed et al., 2013a). These properties prohibit the use of traditional optimization techniques and often require the use of meta-heuristics like MOEAs. This dissertation contributed to a rigorous assessment of the capability of MOEAs to support the resulting high dimensional stochastic control problem for multi-purpose reservoir operations. Several metrics were used to test efficiency, reliability, and controllability of current MOEAs. The overall findings demonstrated that MOEAs that have adopted 1) auto-adaptivity to ensure a diversified exploration of the decision space, 2) mechanisms to keep a stable and bounded search and 3) asynchronous search mechanisms have high potential on reliably optimizing multi-purpose control. The Borg MOEA is demonstrated to have high controllability, improving the approximation of the Pareto by increasing the number of function evaluations (NFE). This opens avenues to test different parallelization schemes and remove the NFE constraint by exploiting high performance computing. Parallelization strategies that are equipped with cooperative search mechanisms

allowed targeting high problem difficulty achieving high fidelity across critical tradeoffs for a large set of Monte Carlo runs, providing confidence in the parallel MOEAs capability to handle uncertainty in multi-purpose control. This discovery provided confidence to extend the control problem to capture pressing concerns for the system under higher dimensionality and a noisier representation of objective functions, challenged now by performance constraints; the MOEA was able to find tradeoffs for this higher complexity. This dissertation contributes to advance multi-purpose reservoir policy design by rigorously validating the suitability of multi-objective evolutionary optimization coupled with direct policy search to solve problem structures that better capture the complexity, interests and constraints in multi-purpose reservoir systems.

1.2 Scope and Organization

Chapter 2: The Lower Susquehanna River Basin Case Study

Chapter 2 describes the Conowingo reservoir system in the Lower Susquehanna River Basin (LSRB) which serves as a test case to advance reservoir control. The LSRB embodies a complex multi-objective system due to its competing demands for hydropower production, environmental flows, cooling water for Peach Bottom Nuclear Power Plant, recreational use and water supply for Baltimore, MD and Chester, PA. The capability of meeting these demands is affected by uncertain hydro-climatic conditions in the region, which makes it a suitable test case to evaluate the impacts of uncertainty in reservoir optimization.

Chapter 3: Methodological Components

Chapter 3 provides an overview of shared methodological components utilized in the detailed studies presented in Chapters 4-6. These include optimal control with Evolutionary Many Objective Direct Policy Search (EMODPS) coupled with Multi-objective Evolutionary Optimization using the Borg MOEA, as well as the stochastic synthetic generation to capture hydro-climatic uncertainty in the LSRB system.

Chapter 4: A Diagnostic Assessment of Evolutionary Algorithms for Multi-objective Surface Water Reservoir Control

Chapter 4 contributes a comprehensive diagnostic assessment of the capability of modern MOEAs to support EMODPS using the Conowingo system. The study demonstrated that modern MOEAs equipped with epsilon-dominance archiving and adaptive search are capable of reliably and effectively finding control policies that balance the LSRB's six-objective tradeoffs using EMODPS opening opportunities for the definition of multi-purpose reservoir control policies.

Chapter 5: Balancing Exploration, Uncertainty and Computational Demands in Many Objective Reservoir Optimization

Chapter 5 demonstrates how to overcome the mathematical and computational barriers associated with capturing uncertainties in stochastic multi-objective

reservoir control optimization. The chapter contributes an assessment of state-of-the-art parallel strategies for the Borg MOEA to support EMODPS, also focusing on the LSRB system's multi-sectorial demands. Increasing statistical fidelity of the objective function evaluations falls at the cost of higher computational demands. This study concludes that emerging self-adaptive parallelization schemes exploiting cooperative search populations are crucial in discovering high quality representations of key operational tradeoffs. Such strategies provide a promising new set of tools for effectively balancing exploration, uncertainty, and computational demands for the identification of multi-purpose reservoir policies.

Chapter 6: Adapting Reservoir Operations to Balance Multi-Sectorial Demands during Floods and Droughts

Chapter 6 discovers new control policies for improving the Conowingo reservoir's operations within the LSRB for floods and droughts. The management model for the Conowingo reservoir system was extended to better capture the conflicting stressors of flood protection and high water supply reliability. A risk-averse robust control formulation of the system's objectives coupled with EMODPS is contributed to better capture compromises that balance the LSRB's multi-sectorial demands while minimizing downstream and dam risks by embedding multiple challenging droughts and floods in the process. A compromise policy selected under this formulation helped improve operations during severe floods relative to historical policies. It also showed high expected annual performance across all sectors, and nearly perfect performance for the critical

requirements of minimum environmental flows, dam protection and nuclear cooling water supply reliability. A compromise release policy defined under a broader set of hydrologic conditions has potential in overcoming the need for sudden decisions during unforeseen emergency conditions in the basin, reducing negative impacts across water demands and vulnerable downstream towns.

Chapter 7: Contributions and Future work

Finally, Chapter 7 summarizes the findings of this dissertation and its contributions to decision support for multi-purpose water reservoir control under uncertainty and to the validity of new methodological advances to improve river basin management. Future research directions are highlighted to improve policy design for multi-purpose systems that incorporate better use of information and opportunities to improve operations by defining robust operations under non-stationary hydrology and socio-economic changes.

1.3 Author Contributions and Collaborative Work

Chapter 4: Jazmin Zatarain Salazar conceived the study and led the optimization, data analysis and writing. Patrick Reed supervised the experiments and contributed to the data analysis and writing. Jonathan Herman provided technical feedback on the computational experiments and contributed to analysis and writing. Matteo Giuliani built the original Conowingo reservoir system simulation and contributed to the data analysis and writing. Andrea Castelletti supervised the experiments and contributed to the data analysis and writing.

Chapter 5: Jazmin Zatarain Salazar conceived the study and led the optimization, data analysis and writing. Patrick Reed supervised the experiments and contributed to the data analysis and writing. Julianne Quinn generated the synthetic streamflow used in the analysis and contributed to the analysis and writing. Matteo Giuliani contributed to the data analysis and writing. Andrea Castelletti supervised the experiments and contributed to the data analysis and writing.

Chapter 6: Jazmin Zatarain Salazar conceived the study and led the optimization, data analysis and writing. Patrick Reed supervised the experiments and contributed to the data analysis and writing. Julianne Quinn contributed to the synthetic streamflow generation and provided technical feedback on the computational experiment and analysis.

CHAPTER 2

THE LOWER SUSQUEHANNA RIVER BASIN TEST CASE

2.1 Susquehanna River Basin Context

The Susquehanna River is the largest river in the eastern United States, flowing through Pennsylvania, New York and Maryland (Figure 2.1). The basin drains over a 71,000 km^2 watershed, providing public water supply to nearly 4.1 million users. The Conowingo reservoir, located in the Lower Susquehanna River Basin (LSRB), plays an important role in balancing the multi-sector water demands for the region.

This is an interstate water body with regional water supply pump stations for Chester, PA and Baltimore, MD. The Conowingo dam was completed in 1968 for hydropower production with a total capacity of 548 MW making it the largest privately operated dam in the U.S. The reservoir has a total capacity of 310,000 acre-ft and plays a key thermal regulation role. Reservoir operations at the Conowingo dam regulate 50% of the freshwater flows going into the Chesapeake Bay. The Conowingo embodies a complex multi-objective system due to the competing demands between hydropower production, FERC environmental flow requirements for fish passage during the migratory season, cooling water for Peach Bottom Nuclear Power Plant, recreational use and water supply for Baltimore, MD and Chester, PA. (illustrated in Figure 2.2). To address these demands, the Susquehanna River Basin Commission has historically led computer-aided adaptive management (Sheer and Dehoff, 2009a) to mediate compromises across the systems multi-sector demands. Giuliani et al. (2014a) contributed an explicit analysis of the Lower Susquehanna River Basin (LSRB)

tradeoffs highlighting important potential conflicts between hydropower revenue, recreation, nuclear power cooling water, and environmental flow requirements. The LSRB system is representative of the management challenges faced in reservoir systems worldwide providing an excellent benchmark to advance decision aiding in reservoir management.

Figure 2.1: Map of the Susquehanna River Basin, highlighted in yellow is the Lower Susquehanna River Basin section. The Conowingo hydropower plant and the Muddy Run facility are represented by a large and a small triangle, respectively, located between Pennsylvania and Maryland.

2.2 Conceptual Model of the Conowingo Reservoir

The Conowingo Reservoir is connected to the Muddy Run Pumped Storage Facility, which provides more than 1,600 MW meeting a large portion of the power demands in the basin. The storage facility was installed to take advantage of intra-daily cycles in energy prices. Namely, during off-peak hours water is pumped uphill from Conowingo Reservoir into the Muddy Run Reservoir; this water is released during peak hours to maximize hydropower profit for the combined system. The power house, located in Conowingo, MD, exploits the reduced pricing associated with excess grid capacity during off peak hours to pump water from the Conowingo Reservoir uphill into Muddy Run, the water then relies on gravity-based return flows to Conowingo to take advantage of peak power demand periods. The Conowingo simulation model used in this study is based on the deterministic formulation in Giuliani et al. (2014a), where a dynamic mass balance between Conowingo Reservoir and Muddy Run is modeled over a historical time series of inflows and evaporation rates as well as the Conowingo and Muddy Run Reservoirs releases. Direct rainfall over the reservoir surface can be negligible in relation to flow from upstream contributing areas. Evaporation, in the other hand, is considered in this instance of the problem formulation since the focus is on prolonged summer droughts were the losses are not negligible. The mass balance relationships are described in Equation 2.1:

$$\begin{aligned} s_{t+1}^{CO} &= s_t^{CO} + q_{t+1}^{CO} + q_{t+1}^{CO,L} - r_{t+1}^{CO} - E_{t+1}^{CO} - q_{t+1}^{p} + r_{t+1}^{MR} \\ s_{t+1}^{MR} &= s_t^{MR} + q_{t+1}^{MR} - r_{t+1}^{MR} - E_{t+1}^{MR} + q_{t+1}^{p} \end{aligned} \tag{2.1}$$

where s_t^i is the volume of water stored at each reservoir (i=Conowingo, Muddy Run), q_{t+1}^{CO} is the mainstem measured at the Marietta gauging station USGS 01576000, and $q_{t+1}^{CO,L}$ is lateral inflow to the Conowingo reservoir, respectively, q_{t+1}^{MR} is the inflow to Muddy Run, and q_{t+1}^{p} is the water pumped from Conowingo to Muddy Run. The volume released is given by the release function: $r_{t+1}^i = f(s_t^i, u_t^i, q_{t+1}^i, E_{t+1}^i)$, (Soncini-Sessa et al., 2007), which depends on the storage s_t^i, the release decision u_t^i, the inflow q_{t+1}^i, and the evaporation loss E_{t+1}^i. The release r_{t+1}^{CO} is a vector of 4 releases that supply water to the nuclear power plant, Baltimore, Chester, and downstream through the Conowingo power plant. The time subscript of each variable represents the time instant at which it assumes a deterministic value. The reservoir storage is measured at time t, whereas inflow has subscript $t + 1$, denoting the inflow to the reservoir in the time interval $[t, t + 1)$. The decision time-step is set at 4 hours to balance the need to follow hourly energy prices and to have a time-step sufficiently long to avoid impact by turbine operation mechanics.

The multi-stakeholder objectives for the dam are modeled over the simulation time horizon H of one year (see Figure 2.2), these objectives are captured in more detail in Section 2.3. Figure 2.2 illustrates the conceptual model and objectives for the Conowingo system.

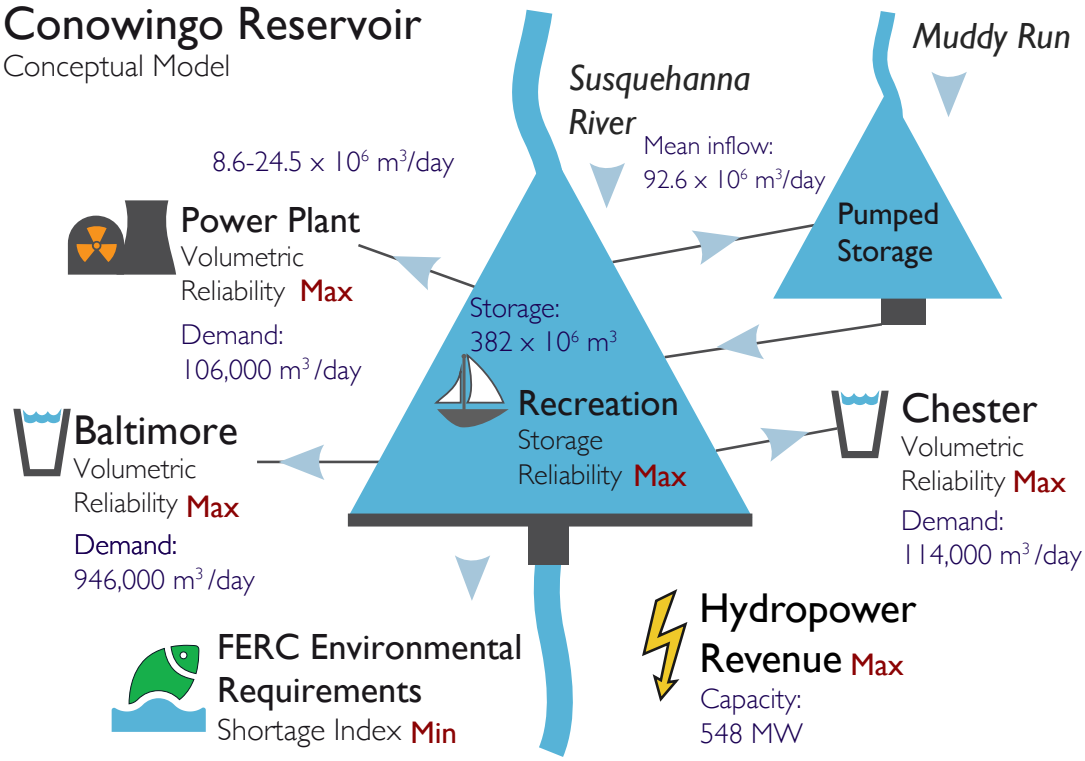

Figure 2.2: Illustration of the six objectives of the Conowingo reservoir. The reservoir has a storage capacity of approximately 310,000 acre-feet and hydropower capacity of 548 MW, with a mean annual inflow from the Susquehanna river of 92.6 10^6 m^3/day. Chester, PA and Baltimore, MD can withdraw a maximum of 114,000 and 946,000 m^3/day respectively, approved by the Susquehanna River Basin Commission. Peach Bottom nuclear Power Station uses water from the Conowingo reservoir for cooling purposes, evaporating up to 10^6,000 m3/day. Additionally, FERC requires a minimum intermittent flow of 8.6 10^6 m^3/day and up to 25.5 10^6 m^3/day for the fish migratory period in the spring (Swartz, 2006).

2.3 Lower Susquehanna River Basin Objectives to be Optimized

2.3.1 Hydropower Revenue

Hydropower revenue (ϕ^{hyd}) is defined as the economic revenue obtained from hydropower production at the Conowingo hydropower plant. Revenue is a function of the hourly energy production (HPt) given in MWh and the hourly energy price (ρ_t) in US \$/MWh defined in Equation 2.2. Energy prices are captured by the seven-hour moving average of the energy price trajectory in the Pennsylvania-New Jersey-Maryland (PJM) energy market (Exelon, 2010).

$$\phi^{hyd} = \sum_{t=1}^{H} (HP_t \cdot \rho_t) \tag{2.2}$$

$$HP_t = \eta g \gamma_w \bar{h}_t q_t^{Turb} \cdot 10^{-6} \tag{2.3}$$

The hourly energy production (MWh) is defined by Equation 2.3, where η is the turbine efficiency, g is the gravitational acceleration (9.81 m/s^2), γ_w is the water density (1000 kg/m^3), $\bar{h}_t$ is the net hydraulic head in meters,that is the reservoir level minus tailwater level and q_t^{Turb} is the turbined flow in m^3/s. This objective is to be maximized in the optimization.

2.3.2 Water Supply Reliability

Water supply reliability ($\Phi^{VR,i}$) to Baltimore, Chester and the nuclear Power Plant needs to be maximized. Each of the water demand objectives is measured as the daily average volumetric reliability defined according to Hashimoto et al. (1982b) in Equation 2.4.

$$\Phi^{VR,i} = \frac{1}{H} \sum_{t=1}^{H} \frac{Y_t^i}{D_t^i} \tag{2.4}$$

where Y_t^i is the daily delivery in m^3, D_t^i is the corresponding daily demand in m^3, $H = 365$ is the management horizon and subscript i represents the water supply to either Baltimore, Chester or to the Nuclear Power Plant.

2.3.3 Storage Reliability for Recreation

Recreation is defined as the storage reliability (Φ^{SR}) in weekends of the tourist season. Given by the relationship between number of weekend days in the tourist season below the target level (TL) and the total number of weekends in the tourist season (N_W). The target level is 32.5 m (106.5 ft) to guarantee recreational activities in the reservoir such as boating.

$$\Phi^{SR} = 1 - \frac{TL}{2N_W} \tag{2.5}$$

2.3.4 Environmental Shortage Index

The environmental shortage index (Φ^{SI}) is to be minimized. Here, it is defined as the daily average shortage index (SI) relative to the Federal Environment Regulatory Commission (FERC) flow requirements (Swartz, 2006). The quadratic function in Equation penalizes larger deficits while allowing small and more frequent shortages (Hashimoto et al., 1982b).

$$\Phi^{SI} = \frac{1}{H} \sum_{t=1}^{H} \left(\frac{max(Z_t - Y_t, 0)}{Z_t} \right)^2 \tag{2.6}$$

CHAPTER 3

METHODOLOGICAL COMPONENTS

3.1 Multi-Objective Evolutionary Optimization

Multi-objective evolutionary algorithms (MOEAs) are population-based heuristic optimization tools inspired by natural selection that abstract the concepts of mutation, crossover and selection to facilitate the discovery of a problem's Pareto set of solutions (i.e., optimal tradeoffs) (Pareto, 1896). The Pareto optimal set is composed of the non-dominated solutions in which performance in any given objective can only be improved by degrading performance in one or more of the remaining objectives. MOEAs are generally designed to solve the following multi-objective vector optimization problem.

$$\begin{aligned} \min_{\mathbf{x}\in\Omega} \quad & F(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots f_D(\mathbf{x})] \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \forall i \in (1, .., M) \\ & h_j(\mathbf{x}) = 0, \forall j \in (1, .., N) \end{aligned}$$

where $f_i(\mathbf{x})$ are the objective functions for D objectives for $\mathbf{x}$ decision variables subject to M inequality constraints and N equality constraints. Historically, the solution of this class of problems required *a priori* assumptions for weighting preferences or schemes that focus on adjusting objective weights for repeated iterations to generate a full representation of the Pareto optimal set of solutions (Cohon and Marks, 1973; Chankong and Haimes, 1983).

This, however, often poses a mix of mathematical requirements for separability, convexity, and deterministic certainty. This approach is not suited for

water resources applications were preferences are not well established *a priori* and highly nonlinear relationships between decisions and outcomes (Reed et al., 2013a) are practically guaranteed due to the stochastic nature of the system. Moreover, it is common to have a mixture of real and discrete decisions, resulting in non-linear and concave decision spaces. Consequently, MOEAs have rapidly grown in popularity and frequency of application due to their population-based search mechanisms; they reduce requirements for providing increased flexibility to represent water systems problems. Through this approach the analyst need not compromise the representation of a complex problem by making simplifications to accommodate for limitations of the solution strategies (Maier et al., 2014a; Nicklow et al., 2010).

Since their early development (Goldberg, 1989; Holland, 1975; Dougherty and Marryott, 1991), evolutionary algorithms have gained wide popularity in a broad suite of applications due to a significant increase in computational power. This has opened the door for larger exploration of solutions to challenging water resources applications (Maier et al., 2014b; Reed et al., 2013a; Nicklow et al., 2010). MOEAs are used to identify the best-known approximation to the Pareto set which are non-dominated by any solution at the end of the search. Formally, to define Pareto optimal solutions, an objective vector $F(\mathbf{x})$ dominates another vector $F(\mathbf{y})$, denoted by $F(\mathbf{x}) \succ F(\mathbf{y})$, iff $f_i(\mathbf{x}) \leq f_i(\mathbf{y})$ $\forall i \in (1, \ldots, D)$ and there exists a $j \in (1, \ldots, D)$ for which $f_j(\mathbf{x}) < f_j(\mathbf{y})$ (Coello et al., 2002).

This reflects point dominance, which may become problematic when dealing with higher dimensional formulations with large numbers of objectives because with a linear growth in objective count each single point is exponentially less likely to be dominated (see proof in Teytaud (2006). Alternatively, ε-

dominance can guarantee stable and bounded convergence and a diverse Pareto front (Laumanns et al., 2002a). ε-dominance requires a level of precision defined for each objective, creating a grid similar to Figure 3.1. Formally, an objective vector $F(\mathbf{x})$ ε-dominates another vector $F(\mathbf{y})$, denoted $F(\mathbf{x}) \succ_\varepsilon F(\mathbf{y})$, iff $(1 - \varepsilon_i) f_i(\mathbf{x}) \leq f_i(\mathbf{y})$ $\forall i \in (1, \ldots, D)$ (Laumanns et al., 2002a).

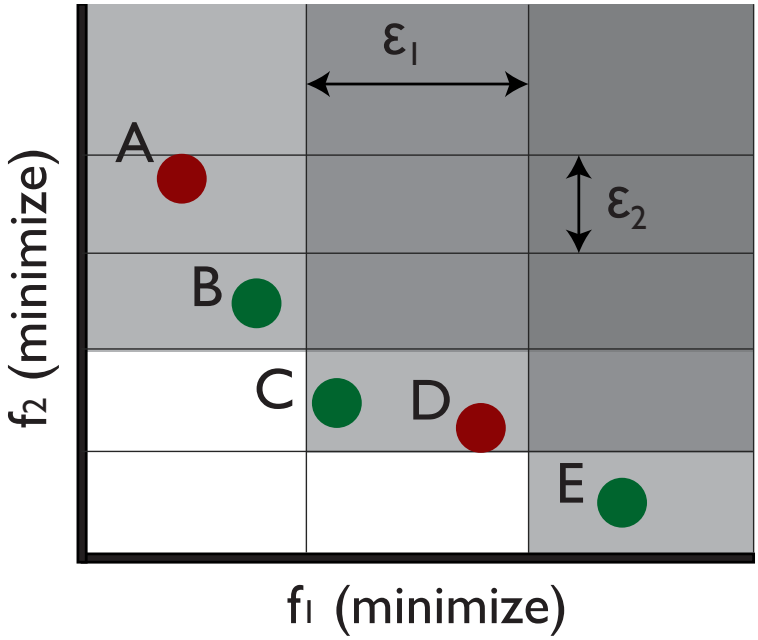

Figure 3.1: Two-objective illustration of ε-dominance indicating precision ε_1 and ε_2 for objective f_1 and objective f_2. Solutions depicted in green circles ε-dominate solutions depicted in red circles, indicating that solution C dominates solution D, and solution B dominates solution A. Figure source: (Kasprzyk et al., 2013a).

3.2 The Borg MOEA

The Borg MOEA (Hadka and Reed, 2013a) was designed to address the optimization challenges associated with formulating and solving many-objective, multimodal real-world applications. The algorithm employs ε-box dominance

archiving (illustrated in Figure 3.1) to maintain convergence and diversity throughout search; ε-progress, to efficiently measure search progression and stagnation; an adaptive population sizing operator based on ε-NSGA-IIs (Kollat and Reed, 2006), time continuation to maintain search diversity and to facilitate escape from local optima; a steady-state, elitist model replacing one solution in the population per iteration enabling its extension in parallel architectures, and multiple recombination operators to enhance search in a wide assortment of problem domains depicted in Figure 3.2.

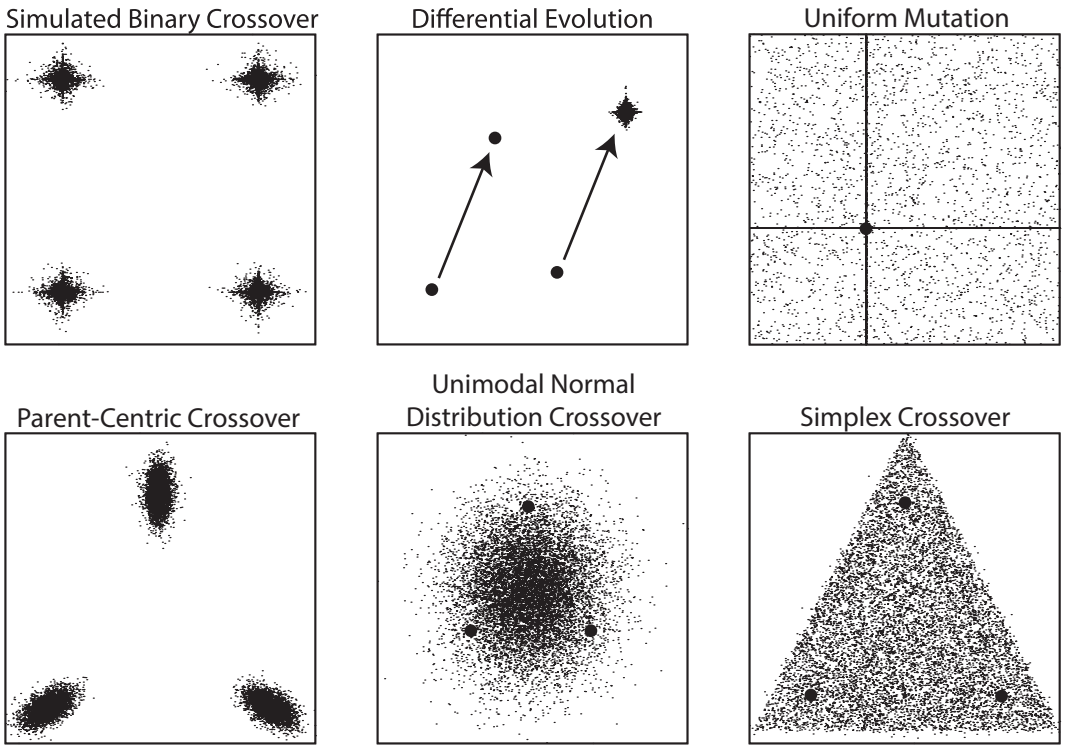

Figure 3.2: Recombination operators used by the Borg MOEA. The illustration shows distributions of new candidate solutions produced by the recombination operators in a hypothetical two-dimensional space of decision variables. The initial parent solutions are indicated by larger circles. Figure source: Hadka and Reed (2013a)

The search operators assimilated by the Borg MOEA include: simulated binary crossover (SBX; Deb and Agrawal (1994)), differential evolution (DE; Storn and Price (1997b), parent-centric crossover (PCX; Deb et al. (2002)), unimodal normal distribution crossover (UNDX; Kita et al. (1999)), simplex crossover (SPX; Tsutsui et al. (1999)), polynomial mutation (PM) and uniform mutation (UM) (Hadka and Reed, 2013a).

Figure 3.3 shows a flowchart of the Borg MOEA main loop. First, one of the recombination operators is selected adaptively based on their probability of successfully finding Pareto approximate solutions. The recombination operator requiring k parents where 1 parent is selected uniformly at random from the archive and the remaining $k-1$ parents are selected from the population through tournament selection, choosing some individuals from the generational population and selecting the ”best” to survive into the next generation. The resulting solutions or offspring are evaluated and can either be placed in the population or the archive. If the solution dominates one or more population members it replaces one of these dominated members randomly. Conversely, if it is dominated by at least one population member, the offspring is not added to the population. Each iteration of this main loop produces one solution. After a certain number of iterations of this loop, ε-progress and the population-to-archive ratio are checked to see if a restart is required. Once the restart has completed, the Borg resumes the main loop until completion. These characteristics have made the Borg MOEA consistently successful for challenging applications in environmental and water resources (Hadka and Reed, 2013a; Reed et al., 2013a; Giuliani et al., 2014a; Quinn et al., 2017a; Zatarain-Salazar et al., 2016; Zatarain Salazar et al., 2017).

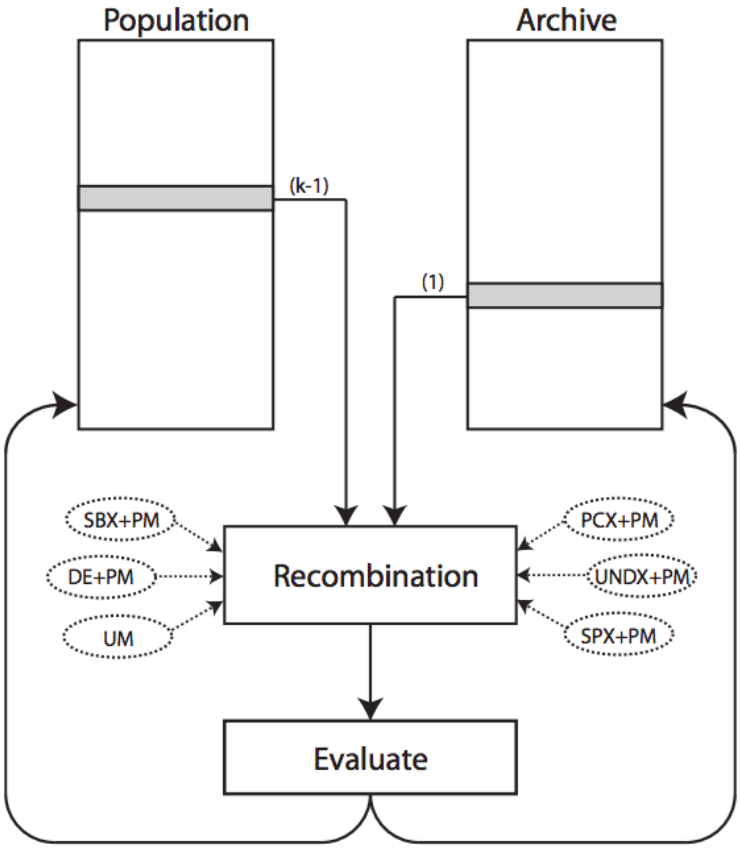

Figure 3.3: The Borg MOEA main loop. Starting with recombination operators using the adaptive operator selection. For a recombination operator requiring k parents, 1 parent is selected uniformly at random from the archive. The remaining $k1$ parents are selected from the population using tournament selection. The solutions resulting are then evaluated for their inclusion in either the population or the archive. Figure source: Hadka and Reed (2013a).

3.2.1 Evolutionary Multiobjective Direct Policy Search (EMODPS)

Evolutionary Multiobjective Direct Policy Search (EMODPS) provides a flexible framework for complex multi-purpose reservoir policy design. Giuliani et al. (2015b) formalized this approach, which features reservoir policy identification, multi-objective evolutionary optimization and visual analytics to characterize the baseline operations and discover the key tradeoffs to provide op-

erators with guidance on balancing a reservoir system's competing demands. Rosenstein and Barto (2001) first introduced direct policy search (DPS) in the general control theory literature. DPS is also known as parameterization-simulation-optimization in the water resources literature (Koutsoyiannis and Economou, 2003) with earlier water resources applications found in Oliveira and Loucks (1997a) and Guariso et al. (1986). EMODPS provides users with flexibility in how to formulate and solve multi-objective reservoir control problems. EMODPS benefits from (1) the simultaneous consideration of heterogeneous forms of objective functions (e.g., minimax and expected value) (Giuliani et al., 2015b; Castelletti et al., 2010), (2) the potential use of exogenous information to condition control decisions, and (3) simulation-based treatment of uncertainties. EMODPS tackles high dimensional reservoir's operational decisions by optimizing the parameters of a control policy. The flexibility and ease of incorporating information into the policy design, broadens the analysis of complex reservoir systems; the systems do not need to be simplified as the methodology can accommodate more objectives and uncertainties without increasing noticeably the problem's difficulty. This process is illustrated in Figure 3.4 for a reservoir optimization problem using as input the time series of inflows into the reservoir's simulation model along with a candidate policy ρ_θ, with initial parameters θ_0. The system trajectories are obtained to calculate the objectives. An MOEA then generates a new set of policy parameters, θ_k for iteration k, that improve the objective values over the maximum specified NFE.

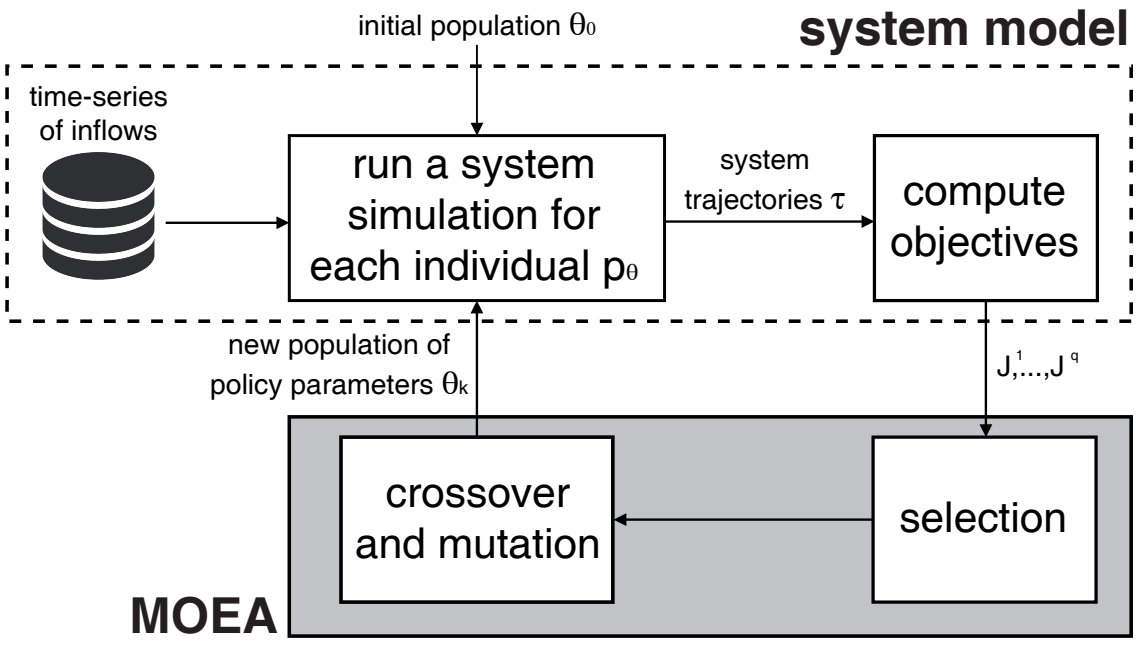

Figure 3.4: Evolutionary Multiobjective Direct Policy Search (EMODPS) used for multi-objective reservoir optimization. First, the time-series of stochastic inputs as well as control policies, ρ_0 defined by initial parameters θ_0 are provided to the systems simulation model, yielding trajectories to compute the objectives. The policies are then refined using an MOEA relative to their performance in the objective functions, this process is repeated for a specified NFE. Figure source: Giuliani et al. (2015b).

3.3 Synthetic Streamflow Generation for the Lower Susquehanna River Basin

This section is drawn from the Supplemental Materials of the following peer-reviewed journal article:

Zatarain Salazar, J., Reed, P. M., Quinn, J. D., Giuliani, M., & Castelletti, A. (2017). Balancing exploration, uncertainty and computational demands in many objective

reservoir optimization. Advances in Water Resources, 109, 196-210.

In this study, operating policies for the Lower Susquehanna River Basin's Conowingo Dam are evaluated over synthetic hydrology to better capture the system's flood and drought extremes since these conditions are rarely observed in the historical record. Consequently, solely considering historical hydrology in the formulation and evaluation of the Conowingo reservoir's operating policies would systematically underestimate the impacts of hydrologic variability and extremes. This supplement provides a description of the methods used to generate synthetic hydrology for the basin (Section 3.3.1 and 3.3.2), as well as a statistical validation of the generator's performance (Section 3.3.3)

Synthetic streamflows at the Marietta gauging station, inflows to Muddy Run, and lateral inflows between Marietta and the Conowingo Dam, as well as evaporation rates over the Conowingo and Muddy Run dams were first generated on a monthly time step using the method of Kirsch et al. (2013). This method is described in Section 3.3.1. Monthly streamflows and evaporation rates at each site were then disaggregated to daily values using the method of Nowak et al. (2010a), described in Section 3.3.3. Since the evaporation rates at the two dams from the OASIS model were identical, only one set was included in the stochastic generation. The rates generated at that site were applied to the other site as well in this study.

3.3.1 Monthly Hydrologic Generation

For a given site, the set of historical streamflows and transformed evaporation rates are denoted as $\mathbf{Q_H} \in \mathbb{R}^{N_H \times T}$ and the set of synthetic streamflows and

transformed evaporation rates as $\mathbf{Q_S} \in \mathbb{R}^{N_S \times T}$, where N_H and N_S are the number of years in the historical and synthetic records, respectively, and T is the number of time steps per year. Here T=12 for 12 months. Evaporation rates were transformed with an exponential transformation because they are approximately normally distributed while the flows are log-normally distributed. For the synthetic generation, the hydrologic variables in $\mathbf{Q_H}$ are log-transformed to yield the matrix $Y_{H_{i,j}} = \ln(Q_{H_{i,j}})$, where i and j are the year and month of the historical record, respectively. The hydrologic variables in $\mathbf{Y_H}$ are then standardized to form the matrix $\mathbf{Z_H} \in \mathbb{R}^{N_H \times T}$ according to Equation 3.1:

$$Z_{H_{i,j}} = \frac{Y_{H_{i,j}} - \hat{\mu_j}}{\hat{\sigma}_j} \tag{3.1}$$

where $\hat{\mu_j}$ and $\hat{\sigma}_j$ are the sample mean and sample standard deviation of the j-th month's log-transformed hydrologic variables, respectively. These variables follow a standard normal distribution: $Z_{H_{i,j}} \sim \mathcal{N}(0, 1)$.

For each site, we generate standard normal synthetic hydrologic variables that reproduce the statistics of $\mathbf{Z_H}$ by first creating a matrix $\mathbf{C} \in \mathbb{R}^{N_S \times T}$ of randomly sampled standard normal hydrologic variables from $\mathbf{Z_H}$. This is done by formulating a random matrix $\mathbf{M} \in \mathbb{R}^{N_S \times T}$ whose elements are independently sampled integers from $(1, 2, ..., N_H)$. Each element of $\mathbf{C}$ is then assigned the value $C_{i,j} = Z_{H_{(M_{i,j}),j}}$, i.e. the elements in each column of $\mathbf{C}$ are randomly sampled standard normal hydrologic variables from the same column (month) of $\mathbf{Z_H}$. In order to preserve the historical cross-site correlation, the same matrix $\mathbf{M}$ is used to generate $\mathbf{C}$ for each site.

Because of the random sampling used to populate $\mathbf{C}$, an additional step is

needed to generate auto-correlated standard normal synthetic hydrologic variables, $\mathbf{Z_S}$. Denoting the historical autocorrelation $\mathbf{P_H}$=corr($\mathbf{Z_H}$), where corr($\mathbf{Z_H}$) is the historical correlation between standardized hydrologic variables in months i and j (columns of $\mathbf{Z_H}$), an upper right triangular matrix, $\mathbf{U}$, can be found using Cholesky decomposition such that $\mathbf{P_H} = \mathbf{U}^\intercal\mathbf{U}$. $\mathbf{Z_S}$ is then generated as $\mathbf{Z_S} = \mathbf{CU}$. Finally, for each site, the auto-correlated synthetic standard normal hydrologic variables $\mathbf{Z_S}$ are converted back to log-space hydrologic variables $\mathbf{Y_S}$ according to $Y_{S_{i,j}} = \hat{\mu}_j + Z_{S_{i,j}}\hat{\sigma}_j$. These are then transformed back to real-space hydrologic variables $\mathbf{Q_S}$ according to $Q_{S_{i,j}} = \exp(Y_{S_{i,j}})$. Finally, the evaporation rates in $\mathbf{Q_S}$ are log-transformed.

While this method reproduces the within-year log-space autocorrelation, it does not preserve year to-year correlation, i.e. concatenating rows of $\mathbf{Q_S}$ to yield a vector of length $N_S \times T$ will yield discontinuities in the autocorrelation from month 12 of one year to month 1 of the next. To resolve this issue, Kirsch et al. (2013) repeat the method described above with a historical matrix $\mathbf{Q'_H} \in \mathbb{R}^{N_{H-1}\times T}$, where each row i of $\mathbf{Q'_H}$ contains historical data from month 7 of year i to month 6 of year *i+1*, removing the first and last 6 months of streamflows from the historical record. $\mathbf{U}'$ is then generated from $\mathbf{Q'_H}$ in the same way as $\mathbf{U}$ is generated from $\mathbf{Q_H}$, while $\mathbf{C}'$ is generated from $\mathbf{C}$ in the same way as $\mathbf{Q'_H}$ is generated from $\mathbf{Q_H}$. As before, $\mathbf{Z'_S}$ is then calculated as $\mathbf{Z'_S} = \mathbf{C}'\mathbf{U}'$. Concatenating the last 6 columns of $\mathbf{Z'_S}$ (months 1-6) beginning from row 1 and the last 6 columns of $\mathbf{Z_S}$ (months 7-12) beginning from row 2 yields a set of synthetic standard normal hydrologic variables that preserve correlation between the last month of the year and the first month of the following year. As before, these are then de-standardized and back-transformed to real space.

3.3.2 Daily Hydrologic Generation

After generating monthly hydrologic variables as described in Section 3.3.1, a nearest-neighbor approach described by Nowak et al. (2010a) is used to disaggregate these hydrologic variables to daily values. The first step in this method is to calculate the k nearest neighbors from the set of historical monthly hydrologic variables for each synthetically-generated month. Nearness is determined by the Euclidean distance, d, in real-space flows and evaporation rates at the 4 sites (equation 3.2):

$$d = \left[\sum_{m=1}^{4} \left((q_S)_m - (q_H)_m \right)^2 \right]^{1/2} \tag{3.2}$$

where $(q_S)_m$ is the real-space synthetic monthly hydrologic variable generated at site m and $(q_H)_m$ is the real-space historical monthly hydrologic variable at site m. For each synthetically-generated hydrologic variable in month j, d is calculated for all historical hydrologic variables in month j. The k-nearest are then sorted from i=1 for the closest to $i = k$ for the furthest, and probabilistically selected for proportionally scaling hydrologic variables in disaggregation. We use the Kernel estimator given by Lall and Sharma (1996) to assign the probability p_n of selecting neighbor n (equation 3.3):

$$p_n = \frac{\frac{1}{n}}{\sum_{i=1}^{k} \frac{1}{i}} \tag{3.3}$$

Following Lall and Sharma (1996) and Nowak et al. (2010a), $k = \left\lfloor N_H^{1/2} \right\rceil$ is used. After a neighbor is selected, the final step in disaggregation is to proportionally scale all of the historical daily hydrologic variables at site m from

the selected neighbor so that they sum to the synthetically generated monthly flow/evaporation rate at site m. For example, if the first day of the month of the selected historical neighbor represented 5% of that month's historical flow, the first day of the month of the synthetic series would represent 5% of that month's synthetically-generated flow.

3.3.3 Verification of Synthetic Streamflow Statistics

As stated previously stated the goal of the synthetic generator is to produce a time series of synthetic hydrologic variables that expand upon those in the historical record while reproducing their statistics. Log-space historical and synthetic probability of exceedance curves of the four hydrologic variables (Figure 3.5) indicate that the first is true, as the synthetic hydrologic variables generate more extreme high and low values. The hydrologic variables also appear unbiased, as this expansion is relatively equal in both directions. Finally, the synthetic probability of exceedance curves also follow the same shape as the historical, indicating that they reproduce the within-year distribution of daily hydrologic variables. Probability of exceedance curves in Figure 3.5 were generated from 1000 years of synthetic hydrologic variables.

To more formally confirm that the synthetic hydrologic variables are unbiased and follow the same distribution as the historical hydrologic variables, we test whether or not the synthetic median and variance of real-space monthly values are statistically different from the historical. The results of these tests are shown in Figure 3.6 for Marietta, which provides most of the system flow. This figure was generated from a 100-member ensemble of synthetic series of

length 100 years, and a bootstrapped ensemble of historical years of the same size and length. Panel a shows boxplots of the real-space historical and synthetic monthly flows, while panels b and c show boxplots of their means and standard deviations, respectively. Because the real-space flows are not normally distributed, the non-parametric Wilcoxon rank-sum test and Levene's test were used to test whether or not the synthetic monthly medians and variances were statistically different from the historical. The p-values associated with these tests are shown in Figures 3.6d and 3.6e, respectively. None of the synthetic medians or variances are statistically different from the historical at a significance level of 0.05.

In addition to verifying that the synthetic generator reproduces the first two moments of the historical monthly hydrologic variables, we also verify that it reproduces both the historical autocorrelation and cross-site correlation at monthly and daily time steps. The results of this analysis, performed on the same synthetic and bootstrapped historical ensembles as for Figure 3.6, are shown in Figures 3.7 and 3.8. Figures 3.7a and 3.7b show the autocorrelation function of historical and synthetic real-space flows at Marietta for up to 12 lags of monthly flows (panel a) and 30 lags of daily flows (panel b). Also shown are 95% confidence intervals on the historical autocorrelations at each lag. The range of autocorrelations generated by the synthetic series expands upon that observed in the historical while remaining within the 95% confidence intervals for all months, suggesting that the historical monthly autocorrelation is well-preserved. On a daily time step, most simulated autocorrelations fall within the 95% confidence intervals for lags up to 15 days, and those falling outside do not represent significant biases.

Figures 3.8a and 3.8b show boxplots of the cross-site correlation in monthly (panel a) and daily (panel b) real-space hydrologic variables for all pairwise combinations of sites. The synthetic generator greatly expands upon the range of cross-site correlations observed in the historical record, both above and below. Table 3.1 lists which sites are included in each numbered pair of Figure 3.8. Wilcoxon rank sum tests (panels c and d) for differences in median monthly and daily correlations indicate that pairwise correlations are only statistically different ($\alpha = 0.5$) between the synthetic and historical series at a monthly time step for site pairs 5 and 6, and at a daily time step for site pairs 1 and 2. Furthermore, biases for these site pairs appear small in panels a and b. In summary, Figures 3.5-3.8 indicate that the streamflow generator is reasonably reproducing historical statistics, while also expanding on the observed record to allow a more thorough stress test of the Lower Susquehanna River Basin's reservoir operating policies.

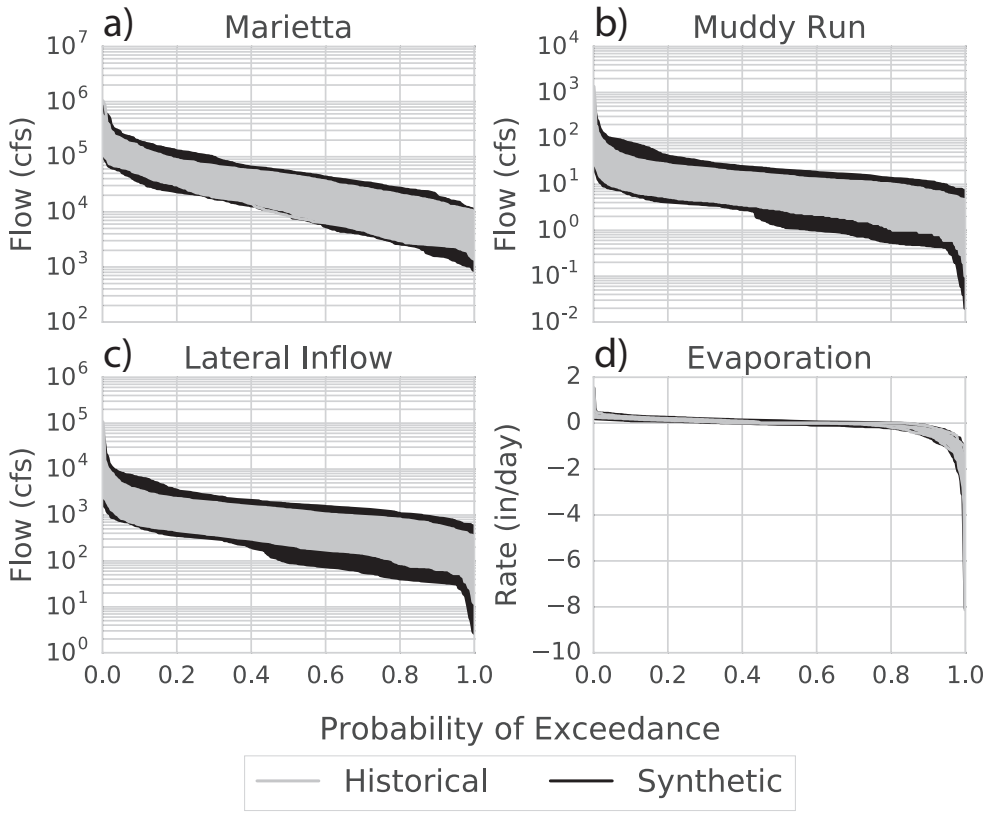

Figure 3.5: Probability of exceedance curves of the historical (gray) and synthetic (black) hydrologic variables in the Lower Susquehanna River Basin. The synthetic hydrologic variables increase the range of values over which the reservoir operating policies are optimized.

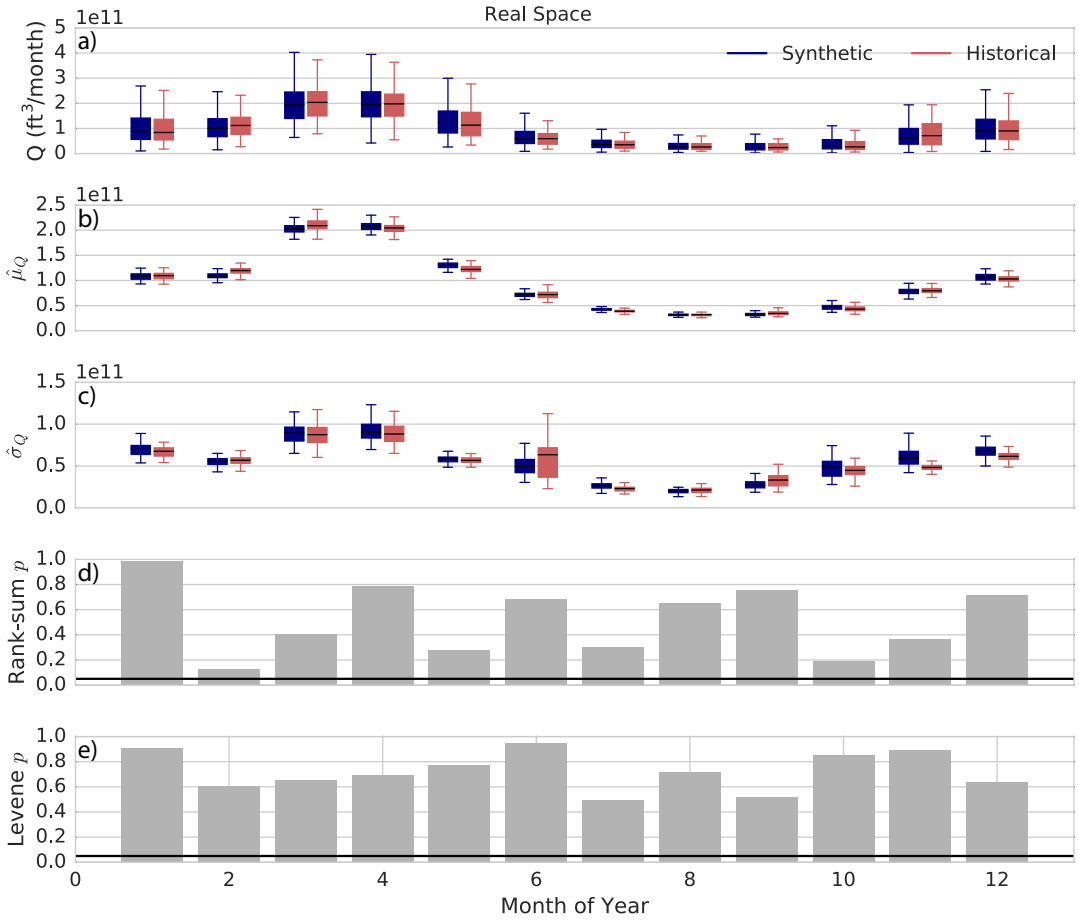

Figure 3.6: Boxplots of the historical (pink) and synthetic (blue) total monthly flows (panel a), mean monthly flows (panel b) and standard deviation of monthly flows (panel c) as well as p-values for differences in median (panel d) and variance (panel e) of monthly flows at Marietta. In the boxplots, a black line is drawn at the median, while the box edges extend to the quartiles and the whiskers to 1.5 times the interquartile range beyond the quartiles. p-values for differences in median were determined by a rank sum test, while those for differences in variance were determined by Levene's test.

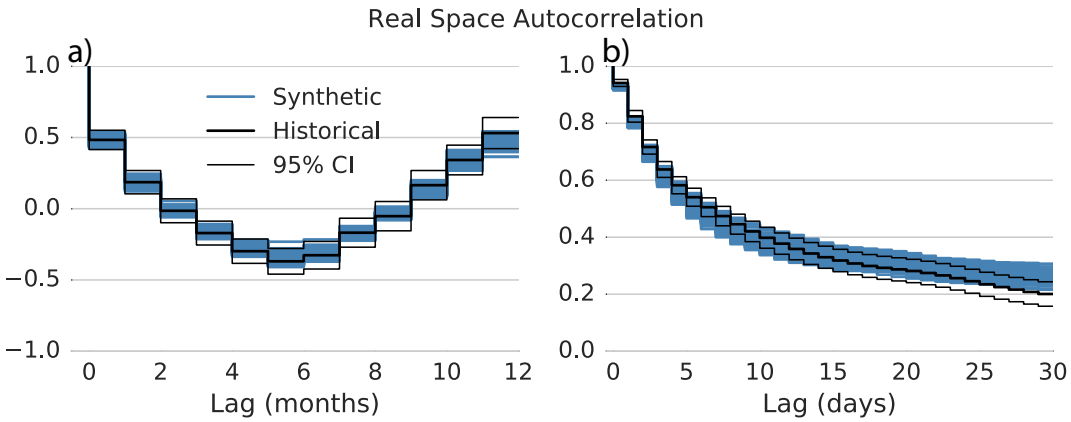

Figure 3.7: Historical (black) and synthetic (blue) monthly (panel a) and daily (panel b) autocorrelation functions for streamflow time series at Marietta. Black lines in panels a and b show both the mean and 95% confidence interval bounds on the historical autocorrelation.

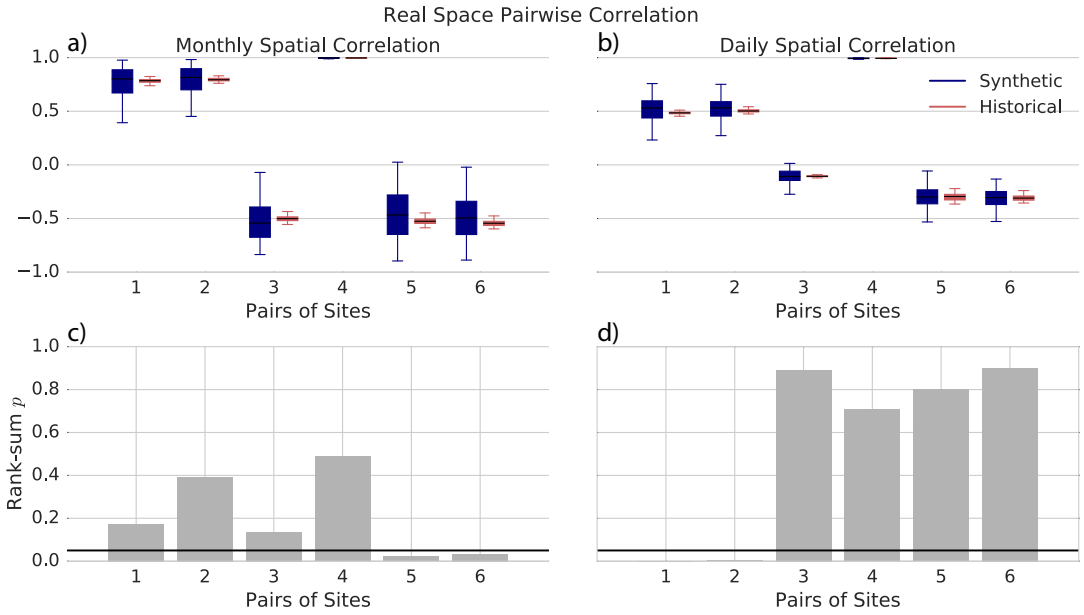

Figure 3.8: Boxplots of pairwise cross-correlations in monthly (panel a) and daily (panel b) historical (pink) and synthetic (blue) hydrologic variables between sites as well as p-values for differences in median (panel c) and variance (panel d). Site pairs are listed in Table 3.1. In the boxplots, a black line is drawn at the median, while the box edges extend to the quartiles and the whiskers to 1.5 times the interquartile range beyond the quartiles.

Table 3.1: Site pairs for the Susquehanna synthetic generation

Pair Number	Sites
1	Marietta and Muddy Run
2	Marietta and Lateral Inflows
3	Marietta and Evaporation
4	Muddy Run and Lateral Inflows
5	Muddy Run and Evaporation
6	Lateral Inflows and Evaporation

CHAPTER 4

A DIAGNOSTIC ASSESSMENT OF EVOLUTIONARY ALGORITHMS FOR MULTI-OBJECTIVE SURFACE WATER RESERVOIR CONTROL

This chapter is drawn from the following peer-reviewed journal article:

Zatarain Salazar, J., Reed, P. M., Herman, J. D., Giuliani, M., & Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. Advances in water resources, 92, 172-185.

Portions of this work were supported by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 as well as the Consejo Nacional de Ciencia y Tecnologia (CONACYT) Fellowship No. 313591. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the US National Science Foundation or CONACYT.

4.1 Abstract

Globally, the pressures of expanding populations, climate change, and increased energy demands are-motivating significant investments in re-operationalizing existing reservoirs or designing operating policies for new ones. These challenges require an understanding of the tradeoffs that emerge across the complex suite of multi-sector demands in river basin systems. This study benchmarks our current capabilities to use Evolutionary Multi-Objective Direct Policy Search (EMODPS), a decision analytic framework in which reservoirs' candidate operating policies are represented using parameterized global approximators (e.g., radial basis functions) then those parameterized functions are opti-

mized using multi-objective evolutionary algorithms to discover the Pareto approximate operating policies. We contribute a comprehensive diagnostic assessment of modern MOEAs' abilities to support EMODPS using the Conowingo reservoir in the Lower Susquehanna River Basin, Pennsylvania, USA. Our diagnostic results highlight that EMODPS can be very challenging for some modern MOEAs and that epsilon dominance, time-continuation, and auto-adaptive search are helpful for attaining high levels of performance. The ϵ-MOEA, the auto-adaptive Borg MOEA, and ϵ-NSGAII all yielded superior results for the six-objective Lower Susquehanna benchmarking test case. The top algorithms show low sensitivity to different MOEA parameterization choices and high algorithmic reliability in attaining consistent results for different random MOEA trials. Overall, EMODPS poses a promising method for discovering key reservoir management tradeoffs; however algorithmic choice remains a key concern for problems of increasing complexity.

4.2 Introduction

Operational water management within river basins worldwide is confronting a challenging combination of growing population pressures, evolving multi-sector demands, and climate change (Pachauri et al., 2014). These challenges are pressing existing and planned hydropower operations to adopt integrated water resources management that takes into account a broad range of social, economic, and environmental issues (Rigg et al., 2009). Efficient multi-purpose reservoir management strategies are critical given the growing risks for flood and drought shocks as well as the need to meet evolving water allocation demands across a complex set of users, e.g., balancing the variability of renewables

or flow maintenance for ecosystem services (Castelletti et al., 2011, 2014a; Kern et al., 2015). However, identifying efficient and balanced reservoir management strategies that meet energy needs while maintaining other key river basin services remains a severe challenge for actual operations.

Reservoir policies need to realistically consider the complex dynamics that typify river basin systems. Consequently, the optimization techniques used in their design need to avoid simplifications that widely discourage their application in real reservoir contexts (Labadie, 2004). Popular operational water management frameworks ranging from classical tools (e.g., alternative dynamic programming (DP) or linear programming(LP) methods) to single-objective heuristics are limited in the breadth of multi-objective formulations that they can resolve (Castelletti et al., 2008, 2014b). Traditionally, these approaches were developed for single objective formulations and only recently they have extended to multi-objective formulations. Yet, they are still limited in their scalability and have not been applied to many-objective formulations (with more than four objectives) (Fleming et al., 2005). The weighting schemes used in traditional multi-criterion implementations of single-objective methods are strongly sensitive to the convexity as well as the separability of the resulting aggregate management objectives (Castelletti et al., 2008, 2012b). These issues pose important limits for formulations with heterogeneous objective functions. For instance, a minimax reliability objective and an expected cost objective may encounter difficulties when integrated into a single weighted function when using a DP framework. The classical approach for appropriately aggregating incommensurable objectives need an *a-priori*, well-specified set of weights (Efstratiadis et al., 2004). Using the terminology of Cohon and Marks (1975a), single objective solution strategies can also be used as generating methods, where a suite of optimiza-

tion runs are executed as the weights for different objectives are varied to attain Pareto optimal solutions (Soncini-Sessa et al., 2007). The Pareto optimal set represents the suite of solutions whose performance in a single objective cannot be improved without degrading their performance in one or more other objectives. Plotting this Pareto optimal set of solutions in a problem's objective space yields the Pareto front, or the geometric representation of the optimal tradeoffs. This scalarization process requires one optimization run for each point that defines a trade-off curve, which is computationally very demanding and often results in poor representations of the Pareto frontier (Castelletti et al., 2013). These limitations make it important to understand the value of algorithms capable of approximating the Pareto front in a single run (e.g., Vamplew et al. (2011); Castelletti et al. (2013); Reed et al. (2013b)). Among these methods, multi-objective evolutionary algorithms (MOEAs) have demonstrated to be capable of discovering high quality representations of complex tradeoffs (Nicklow et al., 2010; Reed et al., 2013b; Maier et al., 2014c; Giuliani et al., 2014a).

Evolutionary Multobjective Direct Policy Search (EMODPS) provides a flexible framework for employing MOEAs in complex multi-purpose reservoir systems. Giuliani et al. (2015a) formalized this approach, which features reservoir policy identification, multi-objective evolutionary optimization and visual analytics to characterize the baseline operations and discover the key operational tradeoffs to provide operators with guidance on balancing a reservoir's competing demands. Rosenstein and Barto (2001) first introduced direct policy search (DPS) in the general control theory literature. DPS is also known as parameterization-simulation-optimization in the water resources literature (Koutsoyiannis and Economou, 2003) with earlier water resources applications found in Guariso et al. (1986) and Oliveira and Loucks (1997a). EMODPS pro-

vides users with flexibility in how to formulate and solve multi-objective reservoir control problems. EMODPS benefits from (1) the simultaneous consideration of heterogeneous forms of objective functions (e.g., minimax and expected value) (Giuliani and Castelletti, 2016), (2) the potential use of exogenous information to condition control decisions, (Giuliani et al., 2015c) and (3) simulation-based treatment of uncertainties in system dynamics or performance (Giuliani et al., 2014a). EMODPS copes with high dimensionality by optimizing directly the parameters of the policy, this is a parsimonious approach that allows to broaden the analysis for complex reservoir systems; the systems do not need to be simplified as the methodology can accommodate more objectives and uncertainties without increasing substantially the problem's difficulty.

Despite these practical advantages, the success of EMODPS is highly dependent on appropriately representing the space of possible operating policies as well as the MOEA's capability to optimize them. The flexibility and accuracy of global approximators to represent alternative operating policies has been assessed in Giuliani et al. (2015a). Although there is a growing number of studies exploring the EMODPS framework, at present no rigorous algorithmic assessments has been completed. The key contribution and focus of this study is to diagnose the difficulty of using MOEAs to support the EMODPS framework using the six-objective Lower Susquehanna test case, descibed in detail in 2.1 and analyze which MOEAs are more suitable for finding the best Pareto approximate set. The Lower Susquehanna test case is challenging due to its large number of conflicting multi-sector demands. Key system demands include hydropower production, urban water supply, recreation and environmental requirements.

4.2.1 Diagnostic Framework

This study implements the comprehensive diagnostic framework illustrated in Figure 4.1 (Hadka and Reed, 2012a; Reed et al., 2013b) to compare the performance of several well known MOEAs. MOEAs are stochastic search tools that use different mating, mutation, selection and archiving parameters. Examples of such parameters can be found in Table 4.2. The default parameterizations for each MOEA are usually defined by finding parameter values that are highly tuned and perform well for specific applications or for test instances; however, as we deal with more complex problems, we encounter less predictable behavior and can no longer assume ideal parameters (Hadka and Reed, 2012a; Reed et al., 2013b). The diagnostic framework used in this study removes this bias by sampling the full feasible parameter space for each evaluated MOEA using Latin Hypercube samples (LHS). Each point drawn from the Latin Hypercube sample in Figure 4.1 represents a full specification of an MOEA's parameters. Furthermore, each parameterization is benchmarked by running the MOEA using multiple randomly generated seeds to account for random effects when generating initial populations and guiding probabilistic search operators. Pareto approximate sets are then computed for each parameter sample. The fitness of the approximation sets is measured through performance metrics that evaluate the solution's convergence, diversity and consistency (discussed in more detail section 4.2.2). The Lower Susquehanna test case has an unknown true Pareto front, hence an approximation of the front was obtained by combining all of the non-dominated solutions attained across all runs of the seven MOEAs' tested. To simplify our nomenclature, in this study the best known Pareto approximation set will be referred to as the reference set, this set is used to calculate the performance metrics.

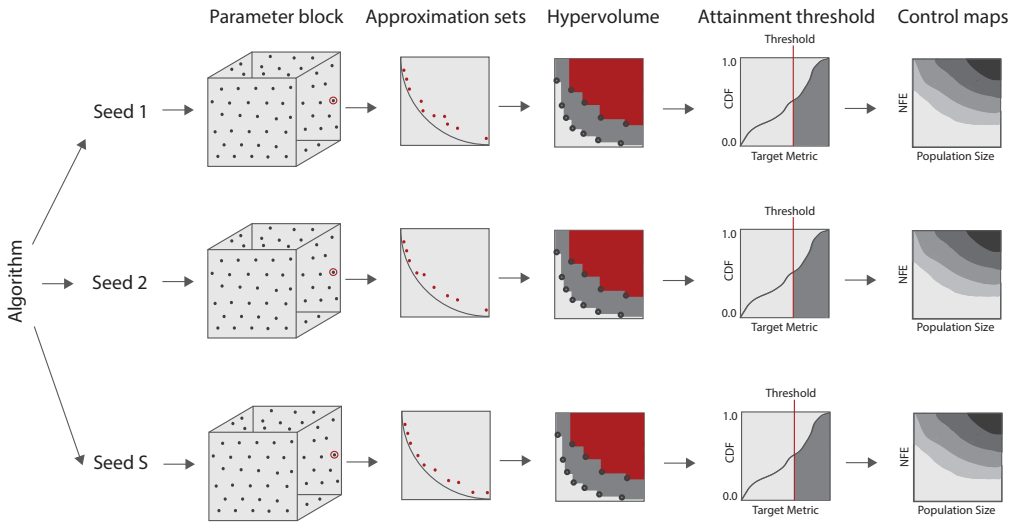

Figure 4.1: Diagnostic assessment framework used to evaluate the performance of each of the seven MOEAs tested in the study (adapted from Reed et al. (2013b)). The parameters for each MOEA are sampled across their full ranges using Lating Hypercube Sampling. Each MOEA parameter set is run in replicate for 30 random seeds to account for random seed effects. The approximation set for each MOEA parameterization is assessed through metrics measuring the convergence, consistency and diversity of approximation sets. Probabilistic assessments of metric attainment are then used to assess the effectiveness and reliability of each algorithm. Visually informed metric goals are then established to construct control maps that capture the efficiency and controllability of each algorithm.

Classification of the performance of an MOEA can be succinctly summarized through its effectiveness, reliability, efficiency, and controllability. Effectiveness measures if an MOEA attains high levels of performance. Reliability captures the variability in performance across parameterizations as well as random seed trials. Efficiency refers to attaining high levels of performance in a minimum number of function evaluations (NFE). Controllability measures the ease-of-use or sensitivity of MOEAs' to their parameterizations. If an algorithm is well suited for a problem, large numbers of its potential input parameter combinations will yield good performance; conversely, if an algorithm is not well suited for a problem, or the problem is particularly difficult, it may be hard to find even a single successful algorithm parameterization. Ideally, an MOEA would generate a high quality Pareto approximation for any combination of its input parameters; this is referred to a large "sweet spot" in its parameter space (Goldberg, 2002b).

Although the MOEA diagnostic framework employed in this study globally samples the algorithms' full feasible parameterization spaces, Reed et al. (2013b) demonstrated that showing performance with respect to the number of function evaluations and population size is sufficient to visualize the ease-of-use and efficiency of the algorithms. Hence, the control maps depicted in Figure 4.1 summarize the full set of sampled parameterizations projected onto the two-dimensional subspace defined by the sampled ranges of NFE and population sizes.

4.2.2 Performance Metrics

The MOEA search performance metrics used in this study are generational distance, ϵ-indicator, and hypervolume. These metrics allow the comparison of the approximation sets by providing an appropriate quantification of proximity, consistency and diversity, respectively (Knowles and Corne, 2002; Zitzler et al., 2003b; Coello Coello, 2007; Hadka and Reed, 2012a). The metrics were calculated relative to the reference set as illustrated in Figure 4.1.

Generational distance (Van Veldhuizen and Lamont, 1998a,b) is the most basic measure of proximity. It consists of the minimum point in the average Euclidean distance vector between solutions in an approximation set and its corresponding nearest solutions in the reference set in the objective space. This is the easiest metric to meet; it only requires a single point to be close to the reference set. The presence of a single point near the reference set will strongly influence the calculation of the minimum Euclidean distance between each of the objective vectors in the approximation set. Generational distance does not account for diversity and if our solution consists of few points that are close to the reference set, this solution has convergence but not diversity, meaning that a good representation of the full set of tradeoffs is not achieved. A large value for this metric implies that an MOEA has failed to find a single solution close to the reference set.

Additive epsilon-indicator (Zitzler et al., 2003b) provides a measure of consistency, where consistency refers to Pareto approximate sets that capture all portions of tradeoffs. The metric is computed as the largest distance that an approximation set needs to be translated in order to dominate the reference set, thus making this metric very sensitive to gaps in tradeoffs. If a Pareto approx-

imate set has gaps, then solutions must be translated a much further distance, dramatically increasing the additive ϵ-indicator metric value (see illustration of this effect in Hadka and Reed (2012a) or Reed et al. (2013b)). A low value for this metric is desired as it measures the worst-case distance from the reference set.

Hypervolume (Zitzler et al., 2003b) provides a measure of diversity and proximity. It quantifies the volume of the objective space dominated by an approximation set; therefore, this metric is to be maximized. In this study, hypervolume was normalized relative to the reference set hypervolume; hence, a value of 1 would mean that the approximation set dominates the same volume as the reference set. Hypervolume is generally the most challenging and comprehensive metric providing insight of an algorithm's convergence and the diversity of their representation of tradeoffs.

4.2.3 Multi-objective Evolutionary Algorithms

Multi-objective evolutionary algorithms are general purpose stochastic search methods simulating natural selection and biological evolution. MOEAs use operators that imitate the processes of mating, mutation and selection to solve multi-objective problem formulations. They start from an initial population of randomly generated solutions, and then seek to iteratively improve this set of solutions using selection, mutation and mating operators. MOEAs population-basis poses important practical advantages that help them deal with challenging mathematical properties such as non-convexity, nonlinearity, stochasticity, and mixtures of continuous as well as discrete decisions. These advantages

are highlighted by (Deb and Gupta, 2006a; Coello Coello, 2007; Deb and Sinha, 2009). Reed et al. (2013b) provide a review of further innovations to MOEAs that dramatically enhanced their performance and usability. These innovations include epsilon-dominance archiving as introduced by Laumanns et al. (2002a) that archives the box non-dominated solutions within a grid with user specified resolution to guarantee convergence and diversity maintenance. Another important MOEA search innovation is termed time continuation, where MOEAs continuously check for search stagnation and introduce new diverse solutions to reinvigorate the search. Most recently, a new class of self-adaptive MOEAs that use feedbacks from their search progress to adapt their strategies have been developed. Self-adaptivity improves the effectiveness, efficiency, reliability, and controllability of the MOEAs (Hadka and Reed, 2012a, 2013a; Reed et al., 2013b), allowing the algorithm to have less dependence upon specific parameter configurations.

Encompassing the above reviewed MOEA innovations, this study benchmarks seven state-of-the- art algorithms that comprise the representative suite of modern tools described below and summarized in Table 4.1. Note this study does not test all ten MOEAs discussed in Reed et al. (2013b); SPEA, IBEA and AMALGAM were left out due to their high algorithmic complexity (Zitzler et al., 2001; Zitzler and Künzli, 2004; Vrugt and Robinson, 2007), which was severely limiting for the Susquehanna test case, requiring high computational demands. Their algorithm wall clock times where several orders of magnitude greater given the same NFE and would have required several months of continuous computational effort.

NSGAII. The Nondominated Sorted Genetic Algorithm II (NSGAII, Deb

Table 4.1: Multi-objective Evolutionary Algorithms (MOEAs) tested in this study

Algorithm	Class	Reference
Borg MOEA	Adaptive multi-operator	Hadka and Reed, 2011
ϵ-NSGAII	Pareto front approximation	Kollat and Reed, 2006
ϵ-MOEA	Pareto front approximation	Deb et al, 2002
OMOPSO	Particle swarm optimization	Sierra and Coello Coello, 2005
GDE3	Differential Evolution	Kukkonen and Lampinen, 2005
MOEA/D	Aggregate functions	Zhang et al, 2009
NSGAII	Baseline	Deb et al, 2000

et al. (2002)) represents a key historical benchmark MOEA and the most widely used algorithm at the time of this study. It features a fast non-dominated sorting procedure that uses Pareto dominance relation to search for the entire Pareto front in a single run. Diversity is preserved using a crowding distance operator to measure how close an individual is to its neighbors. A large average crowding distance results in better diversity. Finally, parents are selected from the population by using tournament selection based on the rank and crowding distance. Key advancements in the NSGAII that have led to its wide use include elitist selection operators (i.e., the best parents are likely to survive) as well as its parameter free crowding distance operator to improve diversity maintenance.

ϵ-MOEA. The Epsilon Dominance Multi-objective Evolutionary Algorithm (ϵ-MOEA, Deb et al. (2003)) is the first instance of an algorithm that actively exploits epsilon dominance archiving as a feedback to search, that is the archive of non-dominated solutions is iteratively updated in the presence of a new solution and the population size is also adapted based on the concept of epsilon-

dominance. It also provides a theoretically guaranteed mechanism for maintaining convergence and diversity maintenance. In multi-dimensional problems convergence maintenance becomes challenging since the number of non-dominated solutions increases very quickly and it becomes difficult to discriminate between solutions; this is also known as dominance resistance. Epsilon-dominance avoids dominance resistance by providing stable and bounded archiving and enables the user to specify the desired precision for each objective allowing to maintain convergence. Epsilon-dominance also provides users with means of improving MOEA efficiency (Kollat and Reed, 2007) by removing the computational burden of seeking Pareto approximate solutions that are not significantly different from a numerical precision perspective. Epsilon dominance requires the user to specify the desired level of precision to identify epsilon non-dominated solutions. ϵ-MOEA is a steady-state algorithm, meaning that only one solution in the population is replaced for each full mating, mutation and selection loop. ϵ- MOEA uses two co-evolving populations: a search population and an archive population. The epsilon-dominance archive is actively exploited in each evolutionary loop where one solution from the population and one from the archive are chosen for generating an offspring solution using the simulated binary crossover (SBX) recombination operator (Deb and Agrawal, 1994) and the polynomial mutation (PM) operator (Deb and Goyal, 1996) to update the archive population.

ϵ-NSGAII. The epsilon dominance NSGAII (ϵ-NSGAII, Kollat (2005); Kollat and Reed (2006)) extends the original NSGAII by including epsilon dominance archiving, adaptive population sizing, and time continuation as part of a limited degree of self adaptive search. The ϵ-NSGAII uses a series of connected runs where small populations are exploited to precondition search with successively

doubled population sizes. Pre-conditioning occurs by injecting current solutions within the epsilon-dominance archive into the initial generations of larger population runs. For example, when an initial smaller population evolves until it is no longer making significant progress, then the population size increases, where 25% of the new population is composed of archived solutions and the remaining 75% are randomly generated (i.e., time continuation of search).

MOEA/D. The Multi-objective Evolutionary Algorithm Based on Decomposition (MOEA/D, Zhang and Li (2007)) exploits the decomposition strategy used in traditional multi-objective optimization. It decomposes a multi-objective optimization problem into several scalar optimization sub-problems and optimizes them simultaneously solving many single-objective Chebyshev decompositions in a single run. Each sub-problem is optimized by only using information from its neighboring sub-problems. The population is composed of the best solution found so far for each subproblem. Only the current solutions to its neighboring subproblem are exploited for optimizing a subproblem in MOEA/D. Since its introduction, the MOEA/D established itself as a benchmark for new MOEAs by winning the 2009 IEEE Congress on Evolutionary Computation (CEC 2009) competition (Reed et al., 2013b).

OMOPSO. Optimized Multi-objective Particle Swarm Optimization Algorithm (OMOPSO, Reyes-Sierra and Coello (2006)) contributed a popular multi-objective extension of the Particle Swarm algorithm, inspired by the movement of organisms in a bird flock or a fish school. OMOPSO incorporates epsilon dominance and uses a crowding factor for the selection of leaders. For each generation and for each particle, a leader is selected. Selection is made by binary tournament based on the crowding value of the leaders. This proposal uses

two external archives: one for storing the leaders currently being used for performing the flight and another one for storing the final solutions. The crowding factor is used to filter out the list of leaders whenever the maximum limit imposed on such list is exceeded. Only the leaders with the best crowding values are retained.

GDE3. The third evolution step of generalized differential evolution (GDE3, Kukkonen and Lampinen (2005)) is the multi-objective extension of the differential evolution algorithm introduced by Storn and Price (1997a) for global optimization. GDE3 starts with an arbitrary number of objectives and constraints. It introduces an adaptive population and non-dominated sorting by pruning non dominated solutions to decrease the population size at the end of each generation, aimed to improve solution diversity and to make it more stable to the population parameter. The diversity maintenance technique is based on a crowding estimation using the nearest neighbors of solutions in a Euclidean sense, and a nearest neighbors search technique. GDE3 uses rotationally invariant operators, producing offspring in any direction relative to the orientation of the fitness landscapes. This is an important characteristic for problems with high dependency among its decision variables (Iorio and Li, 2008; Hadka and Reed, 2013a).

Borg MOEA. Hadka and Reed (2013a) introduced the Borg MOEA, which is not a single algorithm; alternatively, it represents a hyper-heuristic framework (for a review see Burke et al. (2013)), whose search operators are adaptively selected based on the progress being made in solving a problem. The adaptive discovery of key operators is of particular importance for benchmarking how variation operators enhance search for complex many-objective problems. The Borg MOEA assimilates several design principles from existing MOEAs and in-

troduces several novel components. These components include: an ϵ-box dominance archive for maintaining convergence and diversity throughout search; ϵ-progress, which is a computationally efficient measure of search progression and stagnation; an adaptive population sizing operator based on ϵ-NSGAII's (Kollat and Reed, 2005) use of time continuation to maintain search diversity and to facilitate escape from local optima; multiple recombination operators to enhance search in a wide assortment of problem domains; and the steady-state, elitist model of ϵ-MOEA (Deb et al., 2003) which can be easily extended for use on parallel computing (Reed and Hadka, 2014; Hadka and Reed, 2014).

4.3 Computational Experiment

As described in Section 3.2.1, the EMODPS framework abstracts the Conowingo reservoir's operations using RBF-based simulations of alternative operating policies. The experiments are based on the historical formulation in Giuliani et al. (2014a) which are run over the trajectories of inflows, evaporation, and energy prices of 1999 representing a dry, challenging year. The Pareto-approximate set of policies is composed of RBF functions mapping reservoir level and time into reservoir release decisions. Each radial basis function is defined by its center, radius and weight parameters. For the Lower Susquehanna test case, this problem has a total of 32 parameters that are used to generate four release decisions every four hours for a given operational year. These four release decisions are required for water supply to Baltimore, Chester and the Nuclear Power Plant and for downstream release. The RBF output release decisions are a function of inputs for the time index and reservoir level as illustrated in Figure 4.2. The seven MOEAs evaluated in this study seek to identify

Pareto approximate reservoir policies by finding the RBF's shape parameters that yield the best representation of the tradeoffs across the six objectives of the Conowingo reservoir. For those algorithms exploiting epsilon-box dominance, an epsilon precision must be specified to set the acceptable numerical precision to be used for each objective. The epsilon values are 0.5 for hydropower revenue, 0.05 for each volumetric reliability to Baltimore, Chester and the Nuclear Power Plant, 0.05 for recreational storage reliability and 0.001 for the environmental shortage index as in Giuliani et al. (2014a). The overall best known Pareto approximate set for the Lower Susquehanna test case was attained using consistent epsilon-dominance sorting across all algorithms. This reflects that it is always possible to transform point dominance results (e.g., those from NSGAII) into box-dominance results; the reverse is not true. It is critical to maintain consistent dominance relationships when benchmarking MOEAs (Hadka and Reed, 2012a).

The parameter space of each MOEA is sampled using 100 Latin Hypercube samples (LHS). Each point in the parameter space shown in Figure 4.1, represents a full specification of the algorithm's crossover, selection, and mutation operators, as well as initial population sizes and NFEs. Table 4.2 provides a summary of the parameter ranges for each algorithm including Latin Hypercube samples and random seed replicates used in this study. Each point in the parameter block is replicated for 30 random seeds to account for effects on initial populations and probabilistic search operators. Since the Lower Susquehanna test case has an unknown true Pareto Front, the best known approximation, or reference set, is generated by sorting the non-dominated solutions found across 30 seed runs for 100 parameter samples with maximum number of function evaluations of 200,000 for the seven algorithms tested. The performance met-

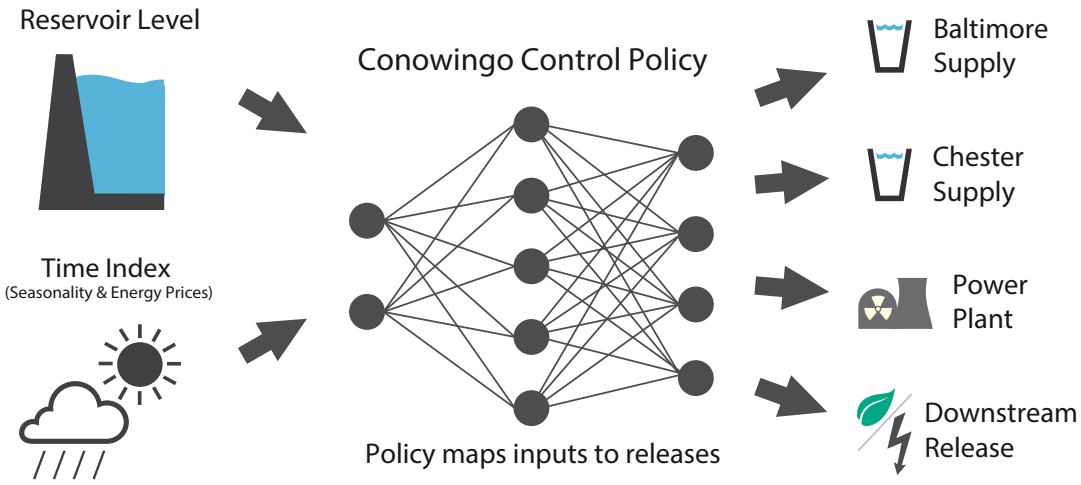

Figure 4.2: Illustration of the control policy represented using radial basis functions. Radial basis functions input reservoir level and time index to define the corresponding release decisions for water supply and downstream release for hydropower production and minimum environmental flow requirements.

rics described in section 4.2.2, are then computed relative to this reference set illustrated in Figure 4.1.

Runtime dynamics for each algorithm were captured in order to test the MOEAs' performance with their default parameterizations. The runtime dynamics provide snapshots of the hypervolume performance attained every one thousand funtion evaluations. The search was extended to 250,000 NFEs to explore if the MOEAs continue to improve their search under their typical use case. Each algorithm was run for 50 random seed trials with their default crossover, mutation and selection operators shown in Table 4.2.

4.4 Results and Discussion

4.4.1 Contributions to Best Known Reference Set

Figure 4.3 represents the reference set of Pareto approximate solutions that compose the tradeoffs for the Conowingo dam potential operating policies. These tradeoffs were obtained across all runs for the seven algorithms tested in this study. The arrows in Figure 4.3 indicate the direction of preference for the environmental shortage index, recreation storage reliability and nuclear power plant volumetric reliability. Although not visible, the ideal solution would be located in back lower corner of the box. Figure 4.3 also provides a measure of the percent of the reference set solutions that were captured by each MOEA within the epsilon precisions specified in section 4.3. Color is used to visualize the portions of the reference Pareto approximate surface contributed by each algorithm. The Borg MOEA and ϵ-MOEA were the largest contributors to the reference set with 48 and 43 percent respectively (shown with large purple and green areas in Figure 4.3). The remaining algorithms made very modest contributions to the reference set where ϵ-NSGAII contributed 4 % of the total solutions; both, NSGAII and MOEA/D contributed with 2%, GDE3 provided 1% contribution. OMOPSO failed to capture any of the reference set solutions. Overall, ϵ-MOEA and the Borg MOEA alone capture 91% of the Lower Susquehanna test case's best known tradeoff solutions.

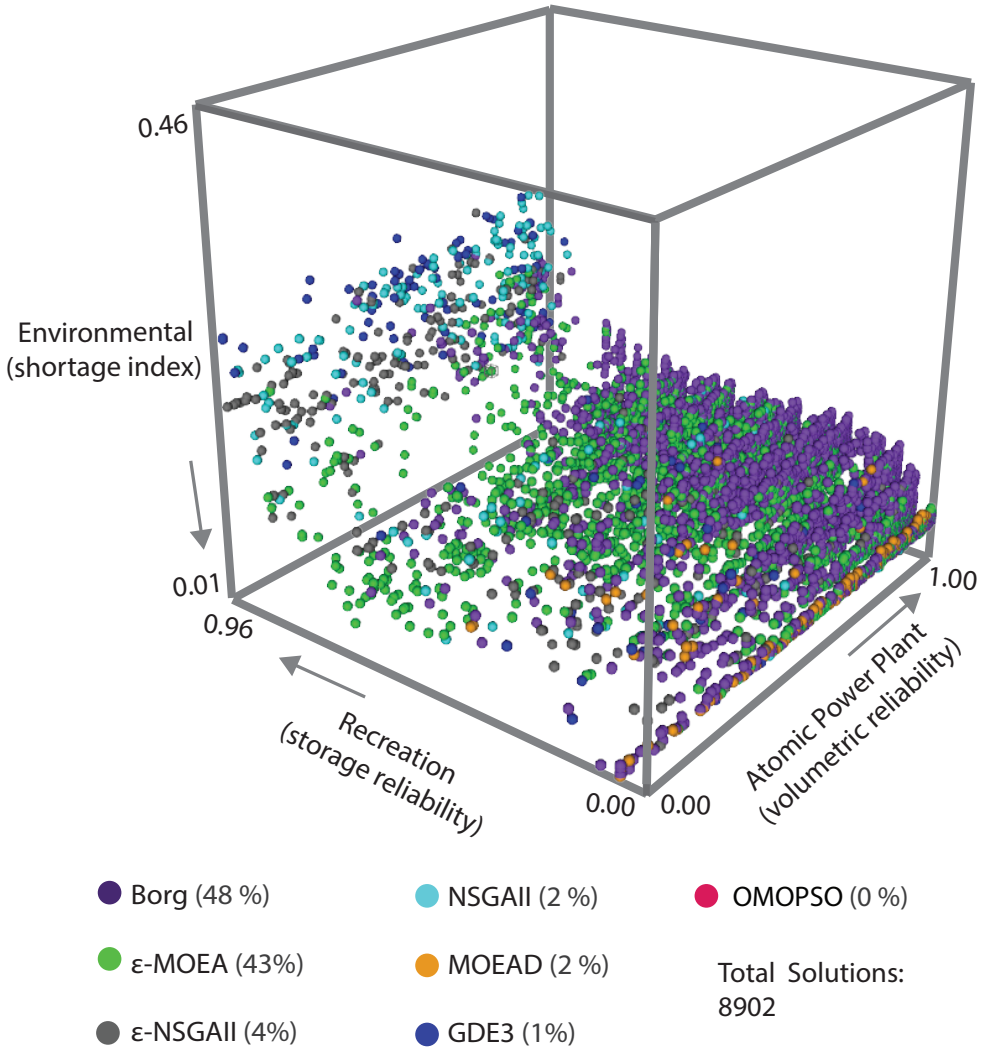

Figure 4.3: The reference set for the Lower Susquehanna test case attained across all MOEA runs. The arrows show the direction of preference for the plotted objectives. The ideal solution would be located in the lower-back corner. The colors represent contributions from each MOEA.

4.4.2 Effectiveness and Reliable Search

To gain a better understanding of the MOEAs effectiveness and reliability, Figure 4.4 provides attainment plots that quantify the probability that a given MOEA attains a certain percent of the best possible metric value. Each of the columns in panels (a)-(c) in Figure 4.4 corresponds to an MOEA. The best overall metric attained in a single seed run for each algorithm is indicated by the black circle in each column. The vertical axis represents the percent of the best metric, and the gray shading indicates the probability of attaining a given percent of the best metric value. Ideal performance would be indicated by a completely black bar with a black circle at the 100% level, designating that a single trial run of an algorithm is both perfectly reliable and effective. Reporting the attainment probability as opposed to only reporting the best overall run, provides a broader context of probabilistic search performance. If an algorithm exhibits outstanding performance in its best single run, but fails to attain high performance consistently, then it has low value to users.

Figure 4.4a illustrates perfect generational distance for all of the MOEAs tested. This is the easiest metric to meet, since it only requires one solution to be close to the reference set. Failing to attain generational distance would indicate that the algorithm failed to find a single solution near the reference set; this metric helps identify complete algorithmic failure. The fact that none of the tested algorithms display abject failure when supporting direct policy search can provide intuition on the difficulty of the problem. Although all of the algorithms attained strong performance in the generational distance metric, this metric is generally only useful for extremely difficult problems where not even a single Pareto approximate solution is identified. However, the nature of a many

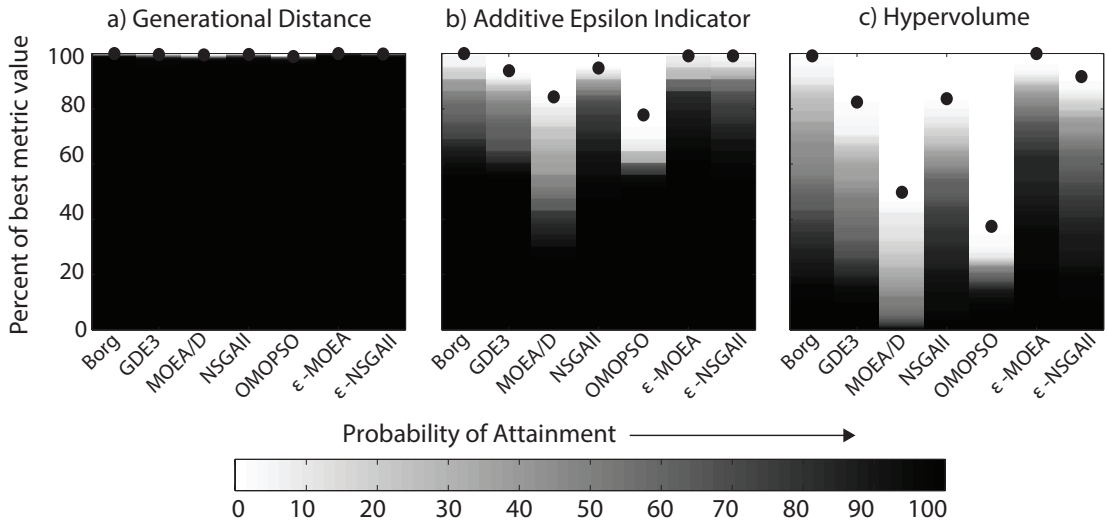

Figure 4.4: The best overall metric value achieved by each MOEA across all runs is designated using black circles. The grayscale shading shows the probability that a single random seed run of an MOEA reaches a given percentage of the best possible metrics values for (a) generational distance, (b) additive epsilon indicator, and (c) hypervolume.

objective problem requires the representation of the full set of tradeoffs, as well as capturing all the portions of the tradeoffs, which this metric fails to provide.

When transitioning to epsilon indicator in Figure 4.4b, all of the algorithms have degraded performance. This is a predictible result since this is a more challenging metric to satisfy and is very sensitive to gaps in the approximation set. Despite the more challenging metric, ϵ-MOEA, the Borg MOEA and

ϵ-NSGAII, are highly effective. The ϵ-MOEA leads the attainment probability for high threshold levels, followed by ϵ-NSGAII and the Borg MOEA. This means that a single seed run of these MOEAs across the tested parameterizations would reliably approximate the Conowingo tradeoffs. It should be noted that in Figure 4.4b, MOEA/D and OMOPSO both show significant degradations in performance.

As expected, Figure 4.4c shows that hypervolume attainment is the most challenging test of performance across all of the algorithms. Hypervolume is typically challenging since it requires high performance for convergence and diversity; in other words, it requires a high quality represention of the full set of tradeoffs. The top performing algorithms are ϵ-MOEA, the Borg MOEA and ϵ-NSGAII. ϵ-MOEA and the Borg MOEA have the best overall metric value; however, their likelihood of attaining high threshold levels decreases. MOEA/D and OMOPSO have the worst performance. OMOPSO's best overall hypervolume achieves only 30% of the best metric value, while MOEA/D achieves 50%. MOEA/D is sensitive to heterogeneous scaling across an application's objectives due to its use of Chebyshev-based aggregations of objectives into a single weighted objective function.

The overall most effective and reliable algorithms based on their best performance and high attainment probability are ϵ-MOEA, the Borg MOEA and ϵ-NSGAII. Any single run of these algorithms would have a good approximation to the best known Pareto policies. The overall results from Figure 4.4 highlight that although the 6-objective formulation of the Lower Susquehanna test represents a very challenging control problem, solving the EMODPS variant of the problem is not that challenging for some of the modern MOEAs.

4.4.3 Controllability and Efficiency

As discussed in section 5.3, we implemented visually informed controllability goals. This refers to the exploitation of visual analytics to determine what corresponding level of the hypervolume performance yields an acceptable representation of the tradeoffs if used in a decision support context. The hypervolume level that represents decision relevant compromises is then used as our target hypervolume. Figure 4.5 illustrates that the 75^{th} percentile hypervolume provides a high quality representation of the reference set. Further refinements of hypervolume performance may not be warranted if they require significantly more computational effort, emphasizing the importance of visualization in MOEA benchmarking as recently noted by Reed and Kollat (2013).

The control maps in Figure 4.6 present two-dimensional projections of the MOEAs performance sampled across their full feasible parameter spaces. They are constructed with population sizes ranging from 10 to 1000 and number of function evaluations ranging from 10,000 NFEs to 200,000 NFEs. These two parameters commonly have a very strong influence on algorithmic performance and computational demands. The color legend provides a measure of the percent of the target hypervolume (see Figure 4.5) captured across 30 random seeds for each of the MOEAs' tested parameterizations. In simpler terms, the results of Figure 4.6 differ from the attainment results of Figure 4.4 by assuming as is typically done in practice that for each of the MOEAs' parameterizations the resultant Pareto approximation set would be developed across 30 independent trial runs. Ideal performance would be represented by an entirely dark blue control map, indicating that an MOEA attained the target hypervolume, across all of its parameterizations (i.e., it is highly controllable). In other words, it indi-

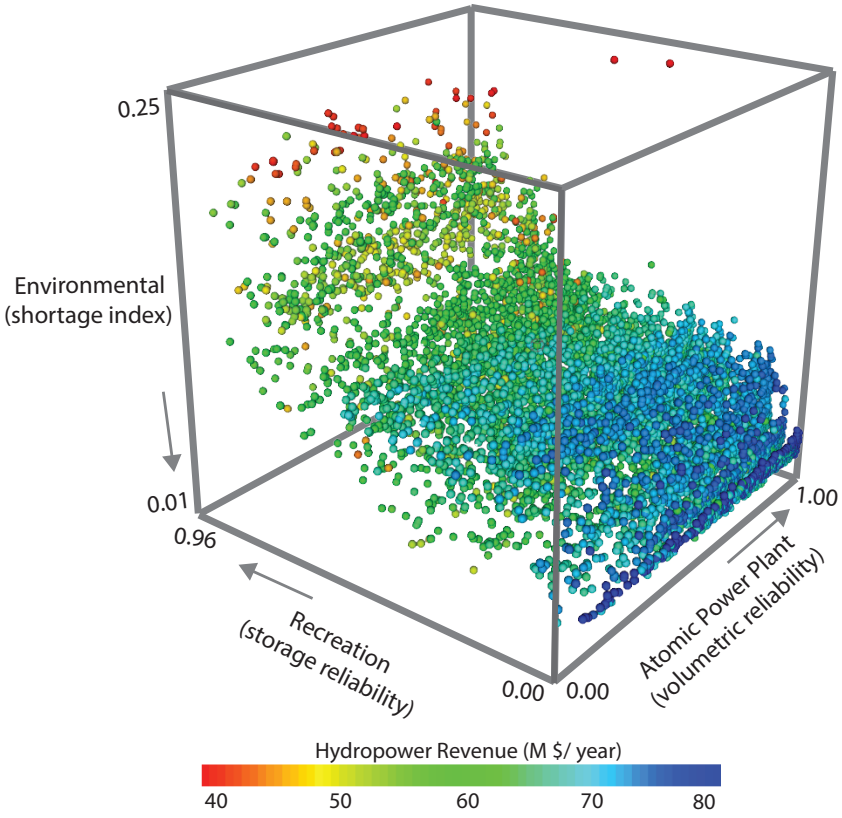

Figure 4.5: Approximation set corresponding to the 75^{th} percentile hypervolume. This performance attainment goal provides a good representation of the Conowingo tradeoffs and defines the target hypervolume for the control maps.

cates that it would be very difficult to make an MOEA fail as a result of how it is parameterized given 30 random seed trials. This would also mean that a good approximation to the Pareto front could be attained with a minimum of NFEs, providing insight of the algorithm's efficiency.

The ϵ-NSGAII and ϵ-MOEA control maps show that these algorithms have a very broad range of effective parameters, indicated by a large dark blue regions in their control maps. These two algorithms are effective for this problem,

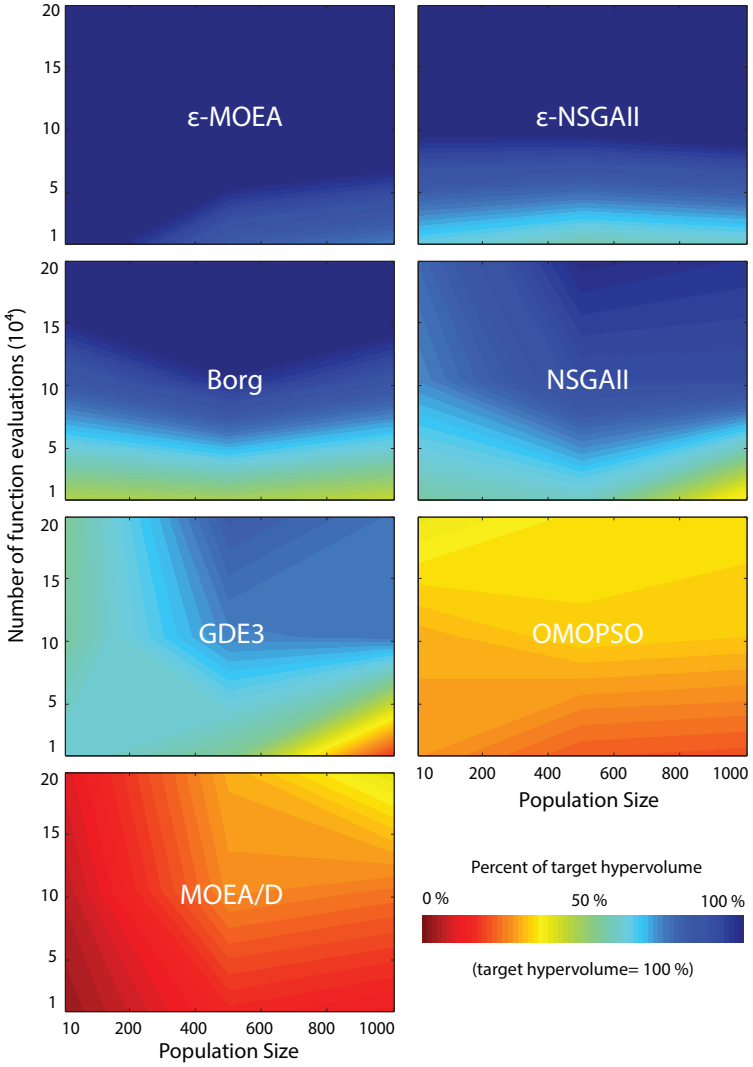

Figure 4.6: Hypervolume performance control maps capturing controllability and efficiency of each MOEA. The color scale represents the percent of the target (75^{th} percentile hypervolume) captured by each local 30-seed approximation set from each tested MOEA parameterization. The control maps are subprojections of the Latin Hypercube samples for each MOEA's full feasible parameter ranges, focusing on the number of function evaluations (NFEs) and population sizes. Ideal performance is shown as dark blue shading indicating that 100% of the target hypervolume is captured by the MOEA, while dark red designates full algorithmic failure.

attaining high levels of performance for a low number of function evaluations, and for a large range of population sizes. The Borg MOEA requires a larger investment of NFEs to maximize its success; it requires a minimum of 100,000 NFEs to consistently attain the target hypervolume level. Consistent with prior findings, the Borg MOEA is generally only sensitive to NFE (Hadka et al., 2012; Reed et al., 2013b). NSGAII, the classical benchmark used in this study, shows a broad zone of high performance; however, it does not attain 100% of the target hypervolume for any of the parameterizations tested. GDE3 requires larger population sizes and increased NFE to improve its performance, requiring more than 100,00 NFE and a population size larger than 400 to attain nearly 75% of the target hypervolume. OMOPSO has poor performance, achieving only 25% to 40% of the target hypervolume for the full range of tested parameters. MOEA/D fully fails for this problem. Although this algorithm has been shown to be very strong in solving test functions, MOEA/D is sensitive to scaling, this makes it less useful for many-objective direct policy search.

The results from Figure 4.6, emphasize the importance of controllable algorithms in water resources applications. Broadly, control maps provide a sense of the sweet spot for an MOEA's parameter space. In other words, they give insight on the sensitivity of an MOEA to its different parameter choices. As we transition to more challenging applications, we need to shift the focus from finding instances of MOEA parameterizations that work for specific applications to MOEAs that support the exploration of challenging problems, and that are capable to yield high quality results regardless of the user-specified parameters. In this regard, the top performing algorithms are the Borg MOEA, ϵ-MOEA, and ϵ-NSGAII, which find broad zones of high performance. Namely, ϵ-MOEA shows high performance for broader parameter ranges than the rest

of the MOEAs; however, the remaining MOEAs show predictible behavior. The Borg MOEA's control map, for example, suggests that it is difficult to make the algorithm fail given sufficient NFE. A similar trend is shown for NSGAII and GDE3; these algorithms have improved results with larger NFE and population sizes. OMOPSO and MOEA/D show the worst overall performance finding Pareto approximate policies for the Conowingo dam. The overall results from Figures 4.4 and 4.6 are consistent in highlighting that the Lower Susquehanna's EMODPS formulation is readily solved by many current MOEAs. The challenge posed in these results is that it would have been very difficult to predict in advance the levels of failure for GDE3, OMOPSO and MOEA/D. When viewed in combination with other recent benchmarking efforts in water resources (Reed et al. (2013b) and Ward et al. (2015)), only the Borg MOEA has consistently performed well across applications. All of the more traditional non-adaptive MOEAs have had mixed success on applications and difficult to predict failure rates. The benchmarking results for EMODPS in this study further highlight that algorithm choice remains a concern.

4.4.4 Runtime Dynamics

To represent the typical use case of running the MOEAs using their default parameterizations, as opposed to sampling the full feasible parameter space, each algorithm was run with their default parameters summarized in Table 4.2. Our typical use case evaluation exploits the MOEAs' absolute hypervolume performance; this metric is clearly demonstrated in Figure 4.4 as being the most challenging given the joint requirements of diversity maintenance and convergence. Figure 4.7 represents the average across 50 random seeds for each algo-

rithm. Their search durations were extended to 250,000 NFE to check for continued progress. MOEA/D failed using its default parameterizations as shown by its horizontal line at 0 hypervolume for the entire run. The runtime dynamics show that MOEA/D's recommended parameterizations are not suited for this problem, even when sampling across a broader range of parameterizations, as shown earlier by the control maps in Figure 4.6, MOEA/D struggles to attain high hypervolume values. NSGAII and GDE3 have moderate improvements within the first 50 thousand NFE; however, they cease to improve for the remainder of their runs. These algorithms are far less effective when using their default parameters. This suggests that a population size of 100 is not sufficient and a larger population may be required to improve their success (as is shown with their control maps in Figure 4.6).

ϵ-MOEA makes rapid progress within the first 100 thousand NFE; however, it exhibits very low improvement throughout the rest of the run. ϵ-NSGAII, outperformed the other algorithms, obtaining the highest hypervolume at the end of the run under default parameterizations, it also showed steep progress within few NFEs, outperforming ϵ-MOEA within 100,000 function evaluations. The Borg MOEA had the second largest hypervolume value at the end of the run; however, its average random seeds show a steady and continued progress throughout the run, indicating that Borg MOEA has potential to improve the search steadily with extended NFE. Both of the top performing algorithms, ϵ-NSGAII and the Borg MOEA, feature time continuation, which enables them to detect stagnation and reinvigorate the search injecting new random solutions. This enables them to continue to improve the search. Additionally, the Borg MOEA's adaptive multi-operator use makes it less sensitive to its parameters and predictably improves performance with increasing NFE.

Table 4.2: Latin hypercube sampling of MOEAs' operators and their associated parameter ranges as well as the MOEAs' default parameterizations.

	Parameter	LHS range	Default	Algorithms
Crossover	SBX rate	0 -1	0.01	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII
	SBX distribution index	0-500	15	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII
	DE crossover rate	0-1	0.1	Borg, GDE3, MOEA/D, NSGAII
	DE step size	0-1	0.5	Borg, GDE3, MOEA/D
	PCX parents	2-10	3	Borg
	PCX offspring	1-10	2	Borg
	PCX eta	0 -1	0.1	Borg
	PCX zeta	0 -1	0.1	Borg
	UNDX parents	2 -10	3	Borg
	UNDX offspring	1 -10	2	Borg
	UNDX eta	0 -1	0.5	Borg
	UNDX zeta	0-1	0.35	Borg
	SPX parents	2-10	3	Borg
	SPX offspring	1-10	2	Borg
	SPX epsilon	0-1	0.5	Borg
Mutation	PM rate	0-1	1	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII, MOEA/D
	PM distribution index	0-500	20	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII, MOEA/D
	UM rate	0-1	1/L	Borg
	Perturbation Index	0-1	NA	OMOPSO
Selection	Neighborhood Size	0-0.2	0.1	MOEA/D
	Delta	0-1	0.9	MOEA/D
	Eta	0-0.02	0.01	MOEA/D
	Archive Size	10-1000	NA	OMOPSO
	Injection Rate	0.1-1	0.25	Borg, ϵ-NSGAII
Population Size		10-1000	100	All algorithms
NFE		10,000 - 200,000	250,000	All algorithms
LHS		100	NA	All algorithms

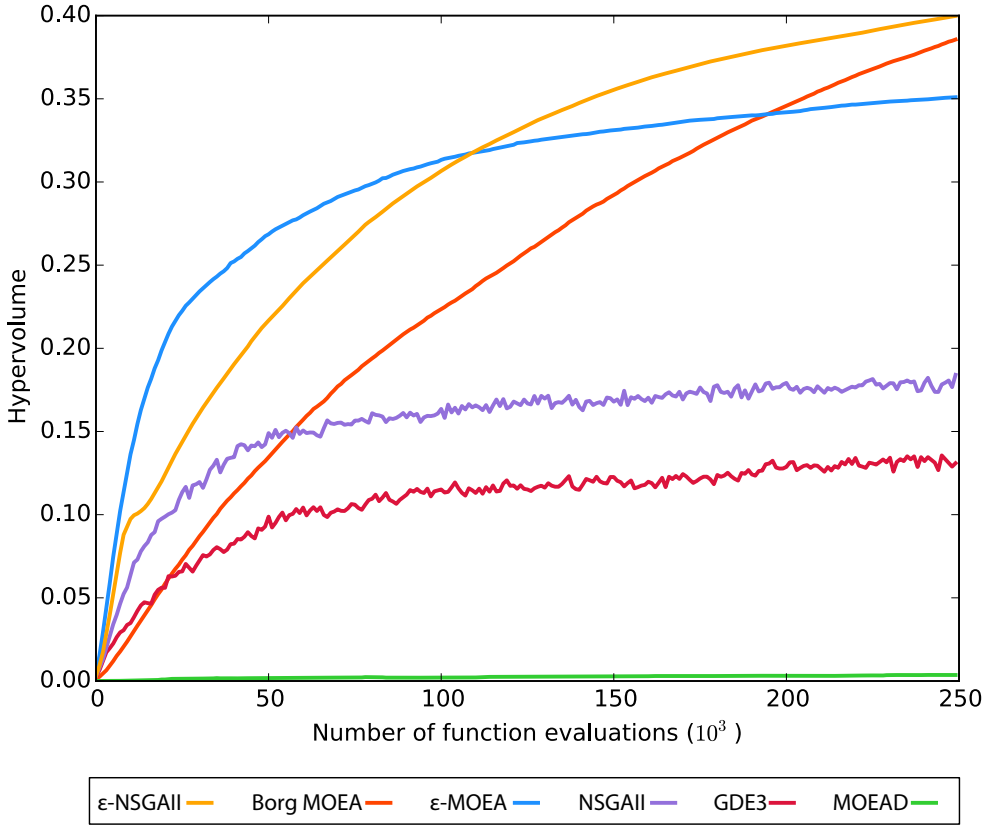

Figure 4.7: Random seed average search dynamics that result when each MOEA solves the Lower Susquehanna test case using their default parameterizations. Each line represents the average hypervolume attained as a function of the number of function evaluations across 50 random seed runs of each MOEA's default parameterizations.

4.5 Conclusions

This study has demonstrated that some modern MOEAs are capable of reliably and effectively finding control policies that balance the Lower Susquehanna test case's six-objective tradeoffs using Evolutionary Multi-objective Direct Policy Search (EMODPS). By acting directly in the policy space, EMODPS enables the evaluation of multiple objectives simultaneously while enhancing users' abil-

ity to exploit simulation models, include complex mixtures of different types of objectives (e.g., expected costs and minimax risks), and incorporate a broader array of system information in reservoir operating policies (e.g., model forecasts, upstream operations, etc.). The potential value of EMODPS increases if it facilitates the transition to complex reservoir control problems.

Understanding our ability to address more complex reservoir management applications using EMODPS requires rigorous assessments of the capabilities of modern MOEAs to support the framework. This study expands on the MOEA diagnostic study by Reed et al. (2013b) by benchmarking the relative efficiency, effectiveness, and reliability of ϵ-MOEA, Borg, ϵ-NSGAII, NSGAII, GDE3, OMOPSO and MOEA/D when solving on a challenging EMODPS application. The diagnostic benchmarking results for these algorithms are based on a six-objective formulation for the management of the Conowingo reservoir located in the Lower Susquehanna River Basin. Overall the ϵ-MOEA, ϵ-NSGAII and the Borg MOEA demonstrated consistently high levels of performance. These three algorithms feature epsilon dominance-archiving which represents a diversity enhancement that also ensures stable and bounded archiving. Additionally, the Borg MOEA and ϵ-NSGAII feature time-continuation with adaptive population sizing to help enhance the search and guarantee continued progress with increased search time. These features help maintain search diversity and facilitate escape from local optima, observed by a continued and stable progress throughout their runtime. MOEA/D and OMOPSO exhibited poor performance; GDE3 and NSGAII also struggled to support direct policy search. This implies that algorithmic performance still remains a concern. Collectively, recent MOEA diagnostic studies demonstrate that only the Borg MOEA has performed consistently well across a wider water applications' suite (Hadka et al.,

2012; Hadka and Reed, 2012a; Reed et al., 2013b; Ward et al., 2015).

As we confront river basin systems with a higher number of reservoirs and increasingly uncertain tradeoffs impacting their operations (e.g., climate change, changes in energy markets, population pressures, ecosystem services, etc.), future EMODPS solution strategies will require MOEAs that are highly scalable and extensible to emerging parallel computing architectures. Future research efforts should focus on extensions that consider broader envelopes of uncertainty to encompass societal challenges and climate change, and also effectively use of information to make policies more adaptive without significantly increasing the computational demands. Reservoirs are very complex and integrative coupled human-natural systems. Effective operations require an understanding of evolving hydro-climatology, conflicting demands, and risks. More research is also needed to bridge the longstanding gap between theoretical analyses of reservoirs and their actual operations.

CHAPTER 5

BALANCING EXPLORATION, UNCERTAINTY AND COMPUTATIONAL DEMANDS IN MANY OBJECTIVE RESERVOIR OPTIMIZATION

This chapter is drawn from the following peer-reviewed journal article:

Zatarain Salazar, J., Reed, P. M., Quinn, J. D., Giuliani, M., & Castelletti, A. (2017). Balancing exploration, uncertainty and computational demands in many objective reservoir optimization. Advances in Water Resources, 109, 196-210.

Portions of this work were supported by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 as well as the Consejo Nacional de Ciencia y Tecnologia (CONACYT) Fellowship No. 313591. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the US National Science Foundation or CONACYT.

5.1 Abstract

Reservoir operations are central to our ability to manage river basin systems serving conflicting multi-sectoral demands under increasingly uncertain futures. These challenges motivate the need for new solution strategies capable of effectively and efficiently discovering the multi-sectoral tradeoffs that are inherent to alternative reservoir operation policies. Evolutionary many-objective direct policy search (EMODPS) is gaining importance for discovering the multi-sectoral tradeoffs associated with reservoir operations due to its capability of addressing multiple objectives and its flexibility in incorporating multiple sources of uncertainties. This simulation-optimization framework has high potential for

addressing the complexities of water resources management, and it can benefit from current advances in parallel computing and meta-heuristics. This study contributes a diagnostic assessment of state-of-the-art parallel strategies for the auto-adaptive Borg Multi Objective Evolutionary Algorithm (MOEA) to support EMODPS. Our analysis focuses on the Lower Susquehanna River Basin (LSRB) system where multiple sectoral demands from hydropower production, urban water supply, recreation and environmental flows need to be balanced. Using EMODPS with different parallel configureations of Borg, we optimize operating policies for the LSRB system over different size ensembles of synthetic streamflows and evaporation rates. As we increase the ensemble size, or Monte Carlo (MC) sample, we increase the statistical fidelity of our objective function evaluations but at the cost of having higher computational demands, making the optimization more difficult. This study demonstrates how to overcome the mathematical and computational barriers associated with capturing uncertainties in stochastic multiobjective reservoir control optimization, where parallel algorithmic search serves to reduce the wall-clock time in discovering high quality representations of key operational tradeoffs. Our results show that emerging self-adaptive parallelization schemes exploiting cooperative search populations are crucial. Such strategies provide a promising new set of tools for effectively balancing exploration, uncertainty, and computational demands when using EMODPS.

5.2 Introduction

Managing river basin systems represents a major global challenge given increasingly uncertain tradeoffs across sectoral uses due to climate change, grow-

ing population pressures, and continued debate over how to sustain environmental services. The computational demands and mathematical difficulty of balancing the inherent multi-sectoral tradeoffs associated with reservoir operations remain a critical research foci (Moss et al., 2016). Emerging computational platforms hold promise for innovating reservoir management by reducing the degree of simplifications and approximations that have traditionally made it difficult to exploit high fidelity simulation models (Giuliani et al., 2015c). As reviewed by Alba et al. (2013), parallel metaheuristics are opening new avenues for simulation-optimization frameworks across a range of application areas. Similarly, recent reviews in the water resources literature (Maier et al., 2014a; Nicklow et al., 2009) highlight that advances in parallel computing platforms and meta-heuristics have gained popularity across a wide range of applications (e.g., water distribution systems, groundwater management, surface water management, etc.).

At present, multi-objective evolutionary algorithms (MOEAs) represent one of the fastest growing areas in the water resources planning and management literature. The popularity of MOEAs lies in their ability to provide an explicit understanding of the systems' tradeoffs (Reed et al., 2013c), which has been a longstanding focus in water management (Cohon and Marks, 1975b; Haimes and Hall, 1977). When working properly, MOEAs exploit population-based search to discover the full set of Pareto approximate solutions in a single run (see review in Coello et al. (2007)). The Pareto-approximate solutions are those for which improvement in one objective can only be achieved by sacrificing performance in one or more other objectives. MOEAs support simulation-optimization applications with challenging mathematical properties such as non-convexity, nonlinearity, stochasticity, mixtures of continuous and discrete

decision variables, multimodality, and stochastic evaluations (Reed et al., 2013c; Coello et al., 2007). It is increasingly more common in MOEA applications to exploit Monte Carlo (MC) simulations to discover robust operating policies (Hamarat et al., 2014; Mortazavi-Naeini et al., 2015; Zhang et al., 2017; Tsoukalas et al., 2016; Müller and Schütze, 2016).

Despite their advantages, the simulation-based search required by MOEAs can still pose high computational costs. Given their stochastic nature, MOEAs need to be run under multiple random seed trials to account for performance differences that may emerge for different randomly generated initial populations or across randomized sequences of their search operators. These trials pose an additional computational demand that can limit the size or scope of applications. Additionally, the growing prevalence of using Monte Carlo simulations to evaluate uncertain objectives also increases computational demands and adds mathematical difficulty to the search problems (Singh and Minsker, 2008; Ward, 2015; Reed et al., 2013c). Several studies have highlighted an inherent tension between the level of approximation in, Monte Carlo-based evaluations of objectives (i.e., small versus large sampling rates) and degree to which an MOEA is able to explore a space (i.e., the number of function evaluations, NFE) (Nicklow et al., 2009; Gopalakrishnan et al., 2003; Singh and Minsker, 2008; Kasprzyk et al., 2013b). Less approximate Monte Carlo evaluations of objective functions are computationally demanding, thus limiting the scope search in a given period of wall-clock time. Alternatively, a faster but highly approximate Monte Carlo evaluation can mislead MOEAs to mistakenly classify some solutions as being superior in all objectives. Such ranking errors can eliminate prematurely important solutions simply as an artifact of noisy evaluations (see detailed discussions in Deb and Gupta (2006b); Beyer and Sendhoff (2007)).

Current innovations in parallel MOEA strategies can help overcome the computational bottleneck posed by expensive Monte Carlo-based function evaluations while also improving the efficiency, effectiveness and reliability of MOEAs when dealing with noisy optimization (Maier et al., 2014a; Reed and Hadka, 2014; Alba et al., 2013; Tang et al., 2007). Efficiency refers to minimizing the computational time to attain solutions. Effectiveness requires high quality Pareto approximation sets. Algorithmic reliability seeks to maintain high levels of efficiency and effectiveness across all random trials. Reed and Hadka (2014) have made significant efforts to assess the scalability of massively parallel runs that demonstrate the potential of MOEAs for large-scale water management applications. We expand their previous efforts by comprehensively diagnosing the ability of two parallel variants of the self-adaptive Borg MOEA to support many objective reservoir optimization under hydro-climatic uncertainty. The master-worker and multi-master worker variants of the Borg MOEA are evaluated in this study for their ability to find Pareto approximate control policies for the Lower Susquehanna River Basin (LSRB). The LSRB system needs to meet several competing water demands for hydropower production, recreation, drinking water distribution for Baltimore, MD and Chester, PA, cooling water for the Peach Bottom Nuclear Power Plant, and federally mandated environmental flows for fish passage. We build our work on earlier efforts by Giuliani et al. (2013) to effectively capture the LSRB's tradeoffs using Evolutionary Multiobjective Direct Policy Search (EMODPS). EMODPS is a simulation-based approach to discover Pareto approximate control policies for multi-purpose reservoir operations. With EMODPS, the control policy is first parameterized using global approximators (i.e., radial basis functions or artificial neural networks) to provide more flexibility to the shape of the control policy. The parameters

are then optimized with respect to the objective functions using parallel Borg MOEA variants. This approach is part of a family of policy-search methods that aim to overcome the computational limitations of traditional reservoir policy optimization methods such as stochastic dynamic programming (SDP) (Guo et al., 2013; Dariane and Momtahen, 2009; Cui and Kuczera, 2005; Koutsoyiannis and Economou, 2003; Oliveira and Loucks, 1997b; Guariso et al., 1986). These limitations can be summarized by the curse of dimensionality (Bellman, 1957), the curse of modeling (Tsitsiklis and Van Roy, 1996; Faber and Stedinger, 2001), and the curse of multiple objectives (Tejada-Guibert et al., 1995), which limit the number of reservoirs whose operations can be optimized, the amount of external information that can be used to condition the decisions (i.e., observations of inflows, precipitation, demand, etc.), and the capability to obtain the full set of Pareto-optimal solutions. This study analyzes the tradeoff between better representing uncertainty in the stochastic optimization and simplifying this representation to shorten the function evaluation time and allow for greater search by optimizing reservoir operations to different size ensembles of stochastic hydrologic inputs. It also demonstrates that cooperative parallel strategies of the Borg MOEA enable high quality representation of tradeoffs as we transition to more challenging formulations of the LSRB to represent the impact of hydro-climatic uncertainties. Showing for the first time that a relatively modest scale cluster system (i.e., hundreds of compute cores), the cooperative Borg MOEA parallelization scheme is able to better cope with the high dimensional LSRB stochastic control problem and its expensive function evaluations while improving the accuracy of its estimated Pareto set within a limited search time, driving the front towards a better approximation with improved convergence speed and reliability.

In this study, the Lower Susquehanna River Basin(LSRB) system described in detail in Chapter 2. The Conowingo system is simulated using a series of stochastically generated streaflows and evaporation rates that better capture variability of extreme flood and drought conditions and compensate for the limited duration of historical observations (see Section 5.2.1 below). The mass balance model is based on the interaction between the Conowingo Reservoir and the Muddy Run Pumped Storage Hydroelectric Facility, which takes advantage of intra-daily cycles in energy prices. The system uses excess power in the grid to pump water into Muddy Run during off-peak hours, and then returns the water by gravity to Conowingo during peak hours (Figure 2.2).

This study is organized as follows: in Section 5.2.1 the synthetic hydrology generation is explained for this particular study, the description of the synthetic generation is detailed in Chapter 3.3. In Section 5.3, the EMODPS implementation is detailed for the LSRB, the parallel Borg MOEA strategies and performance metric used to measure the quality of the Pareto-approximations are also defined in this Section. In section 5.4, the computational experiment is presented. Section 5.5 shows the results of effectiveness, reliability, efficiency and algorithmic speedup of each configuration tested compared against individual MC sampling schemes, the tradeoffs are visualized across all configurations and samples tested.

5.2.1 Stochastic Hydrology

In this study, statistical synthetic hydrology is used to better capture the LSRB system's flood and drought extremes, given that these conditions are rarely ob-

served in the historical streamflow records. Consequently, solely considering historical hydrology in the formulation and evaluation of the Conowingo Reservoir's operating policies would systematically underestimate the impacts of hydrologic variability and extremes (Loucks et al., 2005). The core mathematical benchmarking challenge in this study is to find control policies and their trade-offs that emerge across the uncertain and highly variable streamflow conditions common to the LSRB. Synthetic hydrologic ensembles provide a means of including MC simulation of hydro-climatic uncertainties while maintaining key temporal and spatial statistical traits.

Exploiting the streamflow generator developed by Kirsch et al. (2013), the LSRB synthetic hydrologic ensembles were generated such that they preserve temporal auto-correlation through Cholesky decomposition, and spatial correlation across multiple gauged streamflow points (i.e., inflows to Conowingo and lateral flows to Muddy Run) through a joint resampling scheme of historical flows at multiple sites. This method has been shown to successfully replicate the seasonal correlation structure while producing extreme flow events outside of those observed in the historic record (Kirsch et al., 2012; Herman et al., 2016). These hydrologic variables were then disaggregated to a daily time-step using the K-nearest neighbor procedure (Nowak et al., 2010b). Historical LSRB streamflows from 1932-2001 were used to build the synthetic generator to simulate time-series of the main and lateral streamflow to Conowingo (q_t^{CO}, $q_t^{CO,L}$), as well as the Muddy Run streamflow (q_t^{MR}). Similarly,evaporation rates at Conowingo (e_t^{CO}) and Muddy Run (e_t^{MR}) are simulated, where $e_t^{CO} = e_t^{MR}$. A total of 10,000 annual traces were generated for each of the above-mentioned variables. Figure 5.1 panel (a) shows the annual historical flow duration curves (in blue) versus the stochastically generated flow duration curves (in gray) for

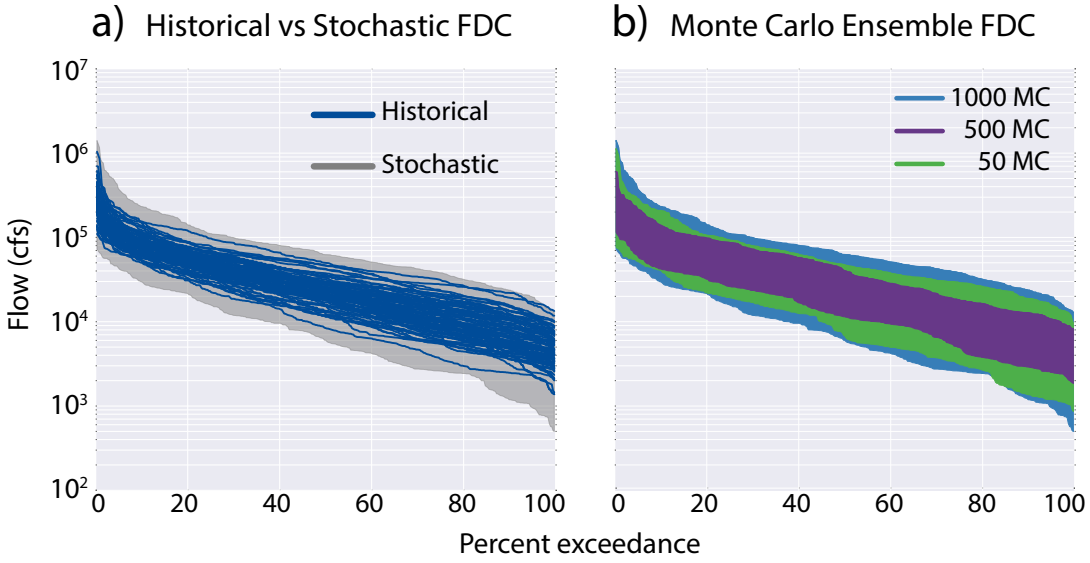

Figure 5.1: Flow duration curves used in the stochastic simulation. Panel (a) shows the annual flow duration curves of the flow at the Marietta gaging station. The observed record (1930-2012) is in blue and the stochastic ensemble is in gray. Panel (b) shows annual flow duration curves sampled from the stochastic ensemble, preserving the wettest and driest years across the ensemble.

q_t^{CO} that provide a broader sample of extreme drought and flood conditions in comparison to the historical flows. Relative to synthetic hydrologic scenarios used in Giuliani et al. (2013), this study added a joint resampling scheme to better capture the spatial correlation between Muddy Run's lateral flows and the Conowingo reservoir's inflows.

As noted in our Introduction, a core challenge and contribution in this study is addressing the complex balance between the computational demands posed by MC-based solution evaluations versus the impacts of highly approximate but fast evaluations. These issues are explored with the LSRB system, including three levels of approximation for the MC evaluations: 50, 500 and 1000 ensem-

ble members. The synthetic hydrologic series used for each MC sampling level were selected from the 10,000 stochastic flows at Marietta, q_t^{CO} shown in Figure 5.1 panel (a). The evaporation rates and the lateral inflow samples were correlated to the Marietta samples. The wettest and driest years at the Conowingo reservoir were included in each ensemble, with the remaining years selected from a one-dimensional Latin hypercube sample of the inverse empirical cumulative distribution of Marietta flows, q_t^{CO}. The flow duration curves of the three approximate MC sampling levels are shown in panel (b) of Figure 5.1.

5.3 Methods

The Conowingo reservoir represents an ideal test case given the array of conflicting multi-sector demands and significant hydro-climatic variability that characterizes the LSRB system. The Conowingo reservoir's operations are explored in this study using the Evolutionary Multi-Objective Direct Policy Search (EMODPS) framework (Giuliani et al., 2015b, 2013). The EMODPS framework provides a generic approach for discovering Pareto approximate reservoir control policies that seek to balance conflicts across competing demands. As discussed in the Introduction, EMODPS is a simulation-based approach for discovering candidate control policies. The operating policies are first parameterized using some form of mathematical functions (e.g., piece-wise linear functions, radial basis functions, neural networks, etc.) and then the parameters of those functions are optimized with respect to the reservoir's operating objectives using MOEAs. The term direct policy search (DPS) was first introduced by Rosenstein and Barto (2001) in the control literature. Independently the same concepts have been termed parameterization-simulation-optimization in the water

resources literature (Koutsoyiannis and Economou, 2003), with early example applications dating back to the 1980s (Oliveira and Loucks, 1997b; Guariso et al., 1986).

One of the most active areas of DPS research lies in extending the framework to multiple objectives. MOEAs are particularly attractive for DPS applications given the potential nonlinearity of candidate control policies, the complexities of capturing uncertainties, and multi-sector tradeoffs. Our previous diagnostic study (Zatarain Salazar et al., 2016) assessed the capability of state-of-the-art MOEAs to support DPS for many-objective reservoir control where the Borg MOEA demonstrated potential for benefiting from parallelization strategies by having continued progress throughout the search, and performing consistently well after a given NFE count. This study contributes a diagnostic study to benchmark different parallelization strategies of the Borg MOEA to discover tradeoffs and support decision making for computationally intensive many-objective reservoir control problems under hydro-climatic uncertainties. Section 5.3.1 presents the control policy formulation. Section 5.3.2 overviews the objective function formulations. Section 5.3.3 presents the Borg MOEA parallel variants. Section 5.3.4 describes the metrics used to test algorithmic efficiency and solution quality.

5.3.1 Control Policy Formulation

In this study, the control policy is parameterized using radial basis functions to map the reservoir level and time index into release decisions. The release from Conowingo (r_{t+1}^{CO}) is actually a vector of 4 releases corresponding to the water

supply for Chester (r^{C}_{t+1}), Baltimore (r^{B}_{t+1}), the Nuclear Power Plant (r^{APP}_{t+1}) and the downstream release for environmental flows and hydropower production (r^{EHP}_{t+1}). The k^{th} policy-prescribed release decision (u^{k}_{t}) in the time interval $[t, t+1)$ is defined as follows:

$$u^{k}_{t} = \sum_{i=1}^{n} w^{k}_{i} \varphi_{i}(z_{t}) \tag{5.1}$$

where n, is the number of RBFs, w^{k}_{i} is the weight of the i^{th} basis function for the k^{th} release, and z_t is the input vector. Each Gaussian RBF (φ) is defined by Equation 5.2:

$$\varphi_{i}(z_{t}) = exp\left[- \sum_{j=1}^{m} \frac{(z_{t} - c_{j,i})^{2}}{b^{2}_{j,i}} \right] \tag{5.2}$$

where m is the number of input variables, $(z_t)_j$ is the value of the j^{th} input variable at time t , and c_{ij} and b_i are the centers and radii, respectively, of the i^{th} RBF for the j^{th} input variable. The parameter vector is defined as $\theta = [c_{i,j}, b_{i,j}, w^{k}_{i}]$, with $i = 1, ..., n$, $j = 1, ..., m$ and $k = 1, ..., 4$ for the 4 release decisions. Here $n = 4$ RBFs are used with $m = 2$ inputs: the time index and the reservoir level. The inputs in z_t are uniformized on $[0, 1]$, while $c_{i,j} \in [-1, 1]$, $b_{i,j} \in [0, 1]$ and $w^{k}_{i} \in [0, 1]$ with $\sum_{i=1}^{n} w^{k}_{i} = 1$.

5.3.2 Formulation of Objectives

The performance of the policy parameters (θ) is evaluated in the objective space by solving the following multi-objective problem:

$$f(\theta^*) = arg \min_{\theta} J(\theta) \tag{5.3}$$

The policy parameters θ^*, are the Pareto approximate decision variables found by the MOEA. The parameters are obtained by simulating the system over a time horizon of one year under the policy $p = \{f(z_t,); t = 0, ..., H - 1\}$ where H is the simulation horizon. The objective functions J are the reservoirs operating objectives, defined by the following objective formulations:

Hydropower Revenue (to be maximized) is defined as the economic revenue obtained from hydropower production at the Conowingo hydropower plant provided by Equation 5.4.

$$J^{hyd} = \sum_{t=1}^{H} (HP_t \cdot \rho_t) \tag{5.4}$$

Revenue is a function of the hourly energy production (HP_t) in MWh and the hourly energy price (ρ_t) in USD/ MWh, defined by a seven hour moving average of the energy price trajectory in the Pennsylvania, New Jersey-Maryland (PJM) energy market. The hourly energy production in MWh is given by Equation 5.5.

$$HP_t = \eta g \gamma_w \bar{h}_t q_t^{Turb} \cdot 10^{-6} \tag{5.5}$$

where η is turbine efficiency, g is the gravitational acceleration (9.81 m/s^2), γ_w is the density of water (1000 kg/m^3), h_t is the net hydraulic head in meters, $q^{Turb} = min(r_t^{EHP}, q^{max})$ is the turbined flow in m^3/s, equal to the minimum of

the downstream release for environmental flows and hydropower production (r_t^{EHP}), and the maximum release through the turbines (q^{max}).

Water Supply Reliability to Baltimore, Chester and the Nuclear Power Plant (to be maximized) is measured by the daily average volumetric reliability in Equation 5.6 (Hashimoto et al., 1982a).

$$J^{VR,i} = \frac{1}{H} \sum_{t=1}^{H} \frac{Y_t^i}{D_t^i} \tag{5.6}$$

where Y_t^i is the daily delivery in m^3, D_t^i is the daily demand in m^3, and the subscript i corresponds to the water supply to either Baltimore, Chester or to the Nuclear Power Plant.

Recreation (to be maximized) is defined as the storage reliability (SR) in weekends of the touristic season described by the relationship between the number of weekend days in the touristic season below the target level (n_F) and the total number of weekends in the touristic season (N_{we}) where the target level is 106.5ft (32.5m) to guarantee boating.

$$J^{SR} = 1 - \frac{n_F}{2N_{we}} \tag{5.7}$$

Environmental Shortage (to be minimized) is defined as the daily average shortage index (SI) relative to the FERC flow requirements described in Equation 5.8. A quadratic function is used in order to penalize larger deficits while allowing small and more frequent shortages (Hashimoto et al., 1982a).

$$J^{SI} = \frac{1}{H}\sum_{t=1}^{H}\left(\frac{max(Z_t - r_t^{EHP}, 0)}{Z_t}\right)^2 \tag{5.8}$$

Z_t and r_t^{EHP} are the FERC flow requirement and the daily environmental release, respectively, both in m^3.

The objectives where evaluated over synthetic realizations of inflow and evaporation rates. A minimax approach formulation was adopted where the objectives are minimized in the worst-case realization. The minimax operator was independently applied for each objective, providing a sense of their lower bound performance. The minimax nature of the objective functions in Equations 5.4- 5.8, make performance evaluations strongly sensitive to MC sampling levels. The three approximation levels for MC sampling explored in this study influence the mathematical difficulty of the optimization as well as the overall computational demands; as the MC sample size increases it becomes harder to find solutions that perform well across the larger ensemble and the computational time per function evaluation increases.

5.3.3 Parallelization Schemes for the Borg MOEA

The serial version of the Borg MOEA (Hadka and Reed, 2013b, 2012b) incorporates key features that have proven to contribute to successful search: (1) the use of epsilon dominance-archiving (Laumanns et al., 2002b); (2) auto-adaptive population sizing and selection; (3) dynamic adaptation and cooperative use of multiple candidate mating and mutation operators; and (4) auto-adaptive triggering of a re-start mechanism that combines the search archive and uniform mutation to diversify search around the best known solutions when progress

stalls. This study compares two parallelization strategies for the Borg MOEA, the simple master-worker scheme and a more recently contributed multi-master scheme (Hadka and Reed, 2015, 2013b). Each of the parallel versions of the Borg MOEA is described in more detail below.

Master-Worker Borg MOEA

The master-worker Borg MOEA illustrated in Figure 5.2, is the simplest parallelization strategy for the algorithm. In the master-worker Borg MOEA only the function evaluations are parallelized; the internal search dynamics of the original serial algorithm are not modified. Consequently, master-worker strategies simply provide the ability to compute more function evaluations in a fixed wall-clock period as there is no interaction between workers (Tang et al., 2007). In a system with C cores, one of the cores is the master and the remaining C-1 cores are labeled as workers. The master node runs the serial MOEA and dispatches the function evaluations to one of the available worker nodes. The master sends the vector of decision variables to an available worker node, the worker node evaluates the problem with the given decision variables, and when finished, sends the evaluated objective values back to the master node.

While the master-worker parallelization scheme will result in the same search dynamics as the serial algorithm, it should be noted that the time savings associated with parallelization can allow for significantly more search given the Borg MOEA utilizes time-continuation of search, (see Goldberg (2002a)). This is especially important for searching water resources applications under uncertainty where MC-based function evaluations can pose significant computational barriers. As mentioned earlier, users must balance the scope of possible search

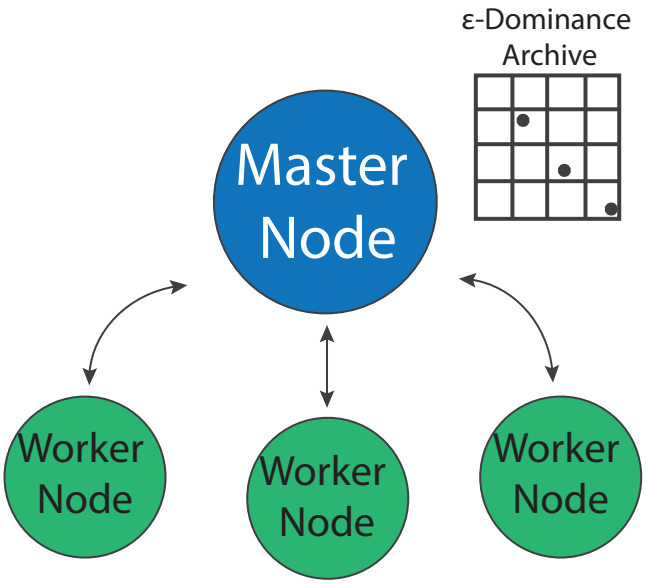

Figure 5.2: Master-worker implementation of the Borg MOEA. The master node performs the main serial loop and dispatches the function evaluations to the available worker nodes which then return the evaluated objective functions to the master node, which keeps track of the epsilon-dominance archive. (Adapted from Hadka and Reed (2015).

(i.e., total NFE) versus the level of approximation used in MC function evaluations (i.e., the number of samples per evaluation). The master-worker Borg MOEA provides a guarantee of scalably increasing the number of function evaluations that can be performed in a given period of unit wall-clock and of maintaining diverse search. Unfortunately, the parallelization scheme does not significantly change the reliability of the algorithm in confronting mathematically challenging problems that induce search failures in the serial algorithm.

Multi-Master Borg MOEA

Beyond simple master-worker schemes, the other most popular implementation of parallel meta-heuristic search has been the multiple search population model, also termed islands or multi-population (see Cantu-Paz (2000)). In this approach, each computing core has a fully functional copy of the search algorithm that interacts with other search populations through cooperative strategies for migrating solutions. The multi-population parallelization scheme has been shown to fundamentally change the meta-heuristics core search dynamics and makes it possible to solve more challenging mathematical problems (see Alba et al. (2013); Tang et al. (2007); Cantu-Paz (2000); Crowl (1994)). Relative to the master-worker, the two largest drawbacks of the multi-population approach are (1) the reduced granularity of its parallelism limits the scope and predictability of maximizing the NFE that can be considered in a given period of wall-clock time due to increased communication and (2) solution sharing or migration schemes add another level of algorithmic complexity that can make it difficult for users to rapidly and reliably maximize search capabilities.

As illustrated in Figure 5.3, Hadka and Reed (2015) developed the multi-master Borg MOEA as a hybrid parallelization scheme that combines the master-worker and multi-population approaches. The multi-master Borg MOEA combines the strengths of both parallelization schemes and generalizes the auto-adaptivity of the algorithm to a highly flexible co-evolutionary framework that can be adapted to a broad range of computing architectures. The multi-master Borg MOEA takes advantage of the cooperative search of distributed populations while maximizing the granularity of its parallelism using worker cores. In its lower bound instantiation, the multi-master Borg

MOEA consists of at least two co-evolving and interacting master-worker instances. Each master independently maintains the self-adaptivity of the serial and single-master implementations of Borg but with the ability to communicate with other masters through a controller in order to maximize performance. The controller introduced by the multi-master Borg MOEA (depicted in Figure 5.3) has two tasks: (1) maintain a global ϵ-dominance archive, and (2) provide guidance to master nodes when they need help. The global ϵ-dominance archive maintains the Pareto optimal solutions discovered by all masters. The controller uses the global ϵ-dominance archive to track the operators that contribute globally Pareto approximate solutions. Each master node periodically sends an update to the controller. This update contains any new Pareto approximate solutions discovered by the master since its last update and the current probability of using each operator.

Since each master node is running an instance of the master-worker Borg MOEA, it includes all of the mechanisms to detect search stagnation and trigger restarts. If these mechanisms are unsuccessful at escaping the local optima, the master node notifies the controller that it needs help. Once receiving the help request, the controller seeds the master with the global ϵ-dominance archive and global operator probabilities that have been successful in contributing new Pareto approximate solutions so far. Upon receiving this guidance from the controller, the master updates its internal state and triggers a restart.

Additionally, the multi-master implementation also features a different style of initialization from the serial and master-worker Borg MOEA implementations. Instead of uniformly sampling the decision variables of the Borg MOEA at random from within their bounds to generate the initial population; the master-

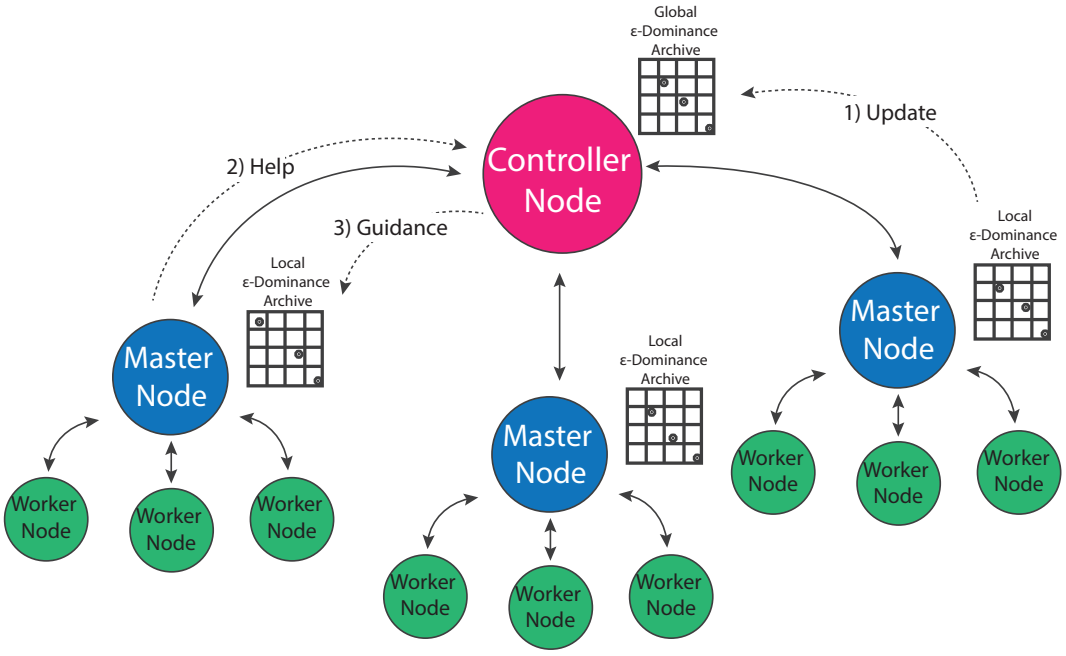

Figure 5.3: Diagram illustrating the multi-master Borg MOEA implementation consisting of two or more master-worker instances. The controller node communicates with the masters in three different ways 1) the masters send periodical updates to the controller of their local epsilon dominance archive; 2) when a master node struggles it sends a help request to the controller, which in turn 3) sends guidance, which includes the global epsilon-dominance archive and operator probabilities (Adapted from Hadka and Reed (2015)).

worker and the multi-master worker implementations used in this study utilize an improved statistical initialization by using a Latin Hypercube sample (LHS) to help ensure that the initial population contains a well-dispersed sampling of the search space.

5.3.4 Measuring the Quality & Speed of Search

An effective MOEA must generate approximation sets that are both proximate (i.e., converge to the true Pareto front) and diverse (i.e., capture the geometry and extents of tradeoffs). Hypervolume (Zitzler et al., 2003a; Knowles and Corne, 2002) is used in this study as a metric that captures both diversity and proximity, to distinguish how well alternative parallel Borg MOEA configurations capture the LSRB's tradeoffs. Figure 5.4 provides an illustrative example of how hypervolume is computed for a 2-objective problem. A reference point is chosen based on the bounds of the approximation set plus an additional delta. This delta ensures the boundary points contribute positive volume to the overall hypervolume. This metric measures the volume of the objective space dominated by an approximation set. A large hypervolume will correspond to approximation sets that dominate more space, indicating high-quality approximation sets (i.e., proximity and diversity).

In addition to using the hypervolume metric to measure solution quality, parallel performance is measured through speedup. Conventionally speedup is defined as the time required to solve an application serially divided by the time required to solve the same application in parallel. However, in the MOEA search context, speedup must be carefully tied to the quality of the search process itself (i.e., hypervolume). This study uses a hypervolume-based speedup analysis to characterize the relative benefits of the multi-master Borg MOEA versus the simpler master-worker variant of the algorithm. A relativistic hypervolume speedup is computed using Equation 5.9. For each level of hypervolume, the algorithmic speedup (S_A) is computed as the ratio of the NFE required by the master-worker (NFE_{mw}) relative to the NFE required per master in the

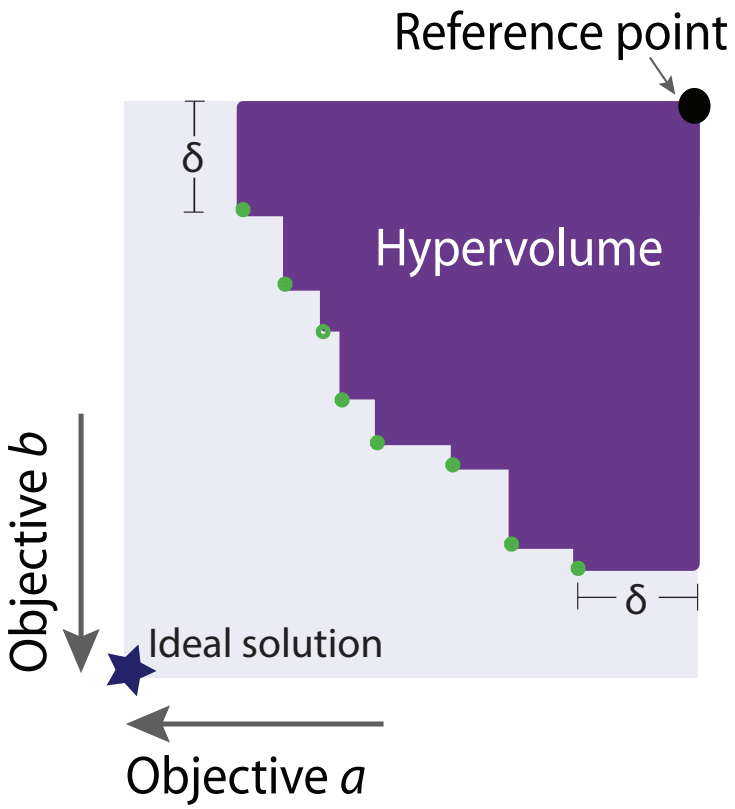

Figure 5.4: Schematic of the hypervolume indicator in a 2D projection. The bounds of the reference approximation set are used to calculate the reference point; this calculation typically adds a delta (δ), so that the boundary points contribute positive hypervolume.

multi-master scheme (NFE_{mm}). In this experiment, the parallel Borg variants are racing to attain high levels of hypervolume performance (HV) within a fixed wall-clock time (T) using the same number of compute cores (C). Any perceived speedups are then the result of improved algorithmic search.

$$S_A = \left(\frac{NFE_{mw}}{NFE_{mm}} \right)_{T,P} \tag{5.9}$$

5.4 Computational Experiment

The computational experiment developed for this study evaluates how well each candidate parallel configuration of the Borg MOEA is able to approximate the LSRB's tradeoffs. Figure 5.5 provides an illustrative overview of the diagnostic evaluation framework.

Three MC sampling levels were used to define the Conowingo reservoir's tradeoffs, calculating objective values across 50, 500 and 1000 independent 1-yr ensemble members shown in panel (a). These simulations were run over different configurations of the master-worker and multi-master Borg MOEA implementations, with core counts ranging from 32 to 256, and master count ranging from 1 to 16 (panel b). The single master-count is equivalent to the master-worker configuration, which serves as the speedup baseline shown in Equation 5.9. Each parallel Borg MOEA configuration was run for 25 random seed trials to account for variability in their initial populations and operator probabilities, each seed was allocated a 10 hour wall-clock time for search. All algorithm configurations were initialized using Latin hypercube samples where each initial Borg MOEA population size was set equal to 100 individuals. The default crossover, mutation and selection parameters of the Borg MOEA used in this study are specified in Table 5.1.

A key difference between the Borg MOEA parallelization schemes is that the master-workers evolutionary processes start from a single initial population while in the multi-master scheme the Latin Hypercube samples are distributed across interacting populations each of which was initiated with 100 members. As discussed in Section 5.3.3 , a key algorithmic innovation that enables the Borg

Table 5.1: Default Parameters of the Borg MOEA

Parameters	Value	Parameters	Value
Injection Rate	0.25	PCX Zeta	0.1
SBX Rate	1	UNDX # of Parents	3
SBX Distribution Index	15	UNDX # of Offspring	2
PM Rate	1/L*	UNDX Eta	0.5
PM Distribution Index	20	UNDX Zeta	0.35
DE Crossover Rate	0.1	SPX # of Parents	3
DE Step Size	0.5	SPX # of Offspring	2
UM Rate	1/L*	SPX Epsilon	0.5
PCX # of Parents	3	Number of islands	1,2,4,8,16
PCX # of Offspring	2	Number of cores	32, 64, 128, 256
PCX Eta	0.1	Wall-clock time	10 hrs.

*L=number of decision variables

MOEA to maintain diverse approximations to the Pareto set is the ϵ-dominance archiving. For the Susquehanna objectives, epsilon precision is established, with 0.5 for hydropower revenue, 0.05 for each volumetric reliability to Baltimore, Chester and the Nuclear Power Plant, 0.05 for recreational storage reliability and 0.001 for the environmental shortage index as in Giuliani et al. (2013). The Pareto approximate sets, depicted as a 2D projection in panel (c), were obtained across the 25 seeds for each algorithm configuration to have a sense of the performance of the best approximation set for each configuration. Additionally, individual approximation sets were obtained for each random seed to test the reliability of each configuration. The absolute hypervolume was computed for each approximation set (panel d). Each set had its own reference point as de-

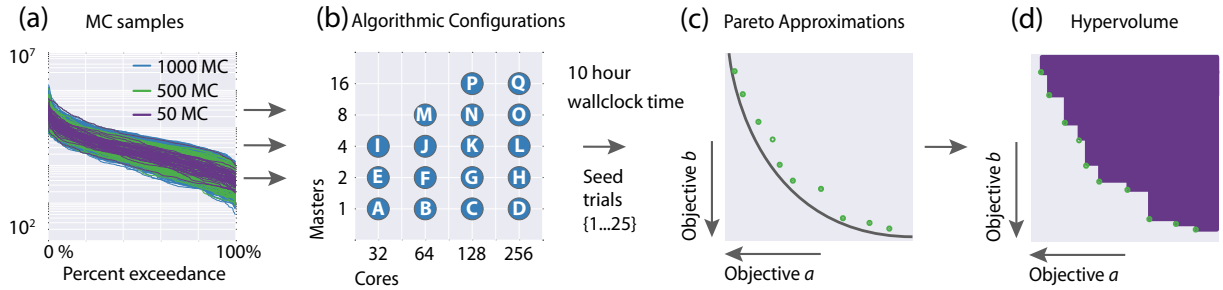

Figure 5.5: Schematic of the experimental setup. Each of the three different problem formulations was run under different algorithmic configurations with different core and master counts which were run for 10 hours. The Pareto approximation sets were collected across masters for each approximation, and finally the quality of the approximation sets was assessed through hypervolume.

scribed in section 5.3.4, which was later normalized to the range [0,1] such that the best possible set was given a hypervolume of 1. Approximation sets with hypervolume near 1 are high-quality, are in close proximity to the best known Pareto front and have captured a diverse representation of the full extent of tradeoffs. Finally, the runtime data was collected every 1000 function evaluations, and hypervolume was computed at each time step to assess runtime performance for every tested Borg MOEA parallel configuration over the 10 hour wall-clock time.

5.5 Results

The diagnostic results explore the balance between exploration, computational demands and the effects of approximation in MC-based evaluations of uncertain

objectives. Section 5.5.1 explores the effectiveness and reliability of each parallel configuration of the Borg MOEA across its tested random seed trials. Effectiveness is explored in terms of the master-worker and multi-master configurations' abilities to attain high quality approximations to the best-known Pareto front (or reference sets). Search reliability refers to the probability that highly effective search is attained. Section 5.5.2 explores the computational efficiency of the parallel Borg MOEA configurations in solving the LSRB Conowingo reservoir test case in terms of the expected NFE required as well as parallel algorithmic speedups. Section 5.5.3 explores the decision relevant consequences of the approximate MC evaluations as well as how choosing a parallel configuration of the Borg MOEA can influence tradeoff analyses.

5.5.1 Effectiveness & Reliability of Search

Figure 5.6 shows the relative hypervolume for the reference sets discovered by each of the algorithmic configurations across their 25 random seed trials. The results were normalized against the best hypervolume attained by each MC sampling level; hence, the parallel Borg MOEA configurations A-Q are only comparable when they exploit the same approximate sampling level (i.e., 50, 500, or 1000 ensemble members). Each circle corresponds to a different parallel configuration of the multi-master Borg, where its location on the x-axis corresponds to its core count and its location on te y-axis to its master count. The color scale represents hypervolume; high performance is depicted in dark blue whereas poor performance is shown in light blue. Panel (a) of Figure 5.6 shows the results for the 50 MC sampling scheme. This sampling scheme strongly reduces computational demands at the cost of using highly approximate function

evaluations to guide the search. Given that in panel (a) NFE are not as strongly limiting, it is not surprising that 12 out of the 17 configurations achieve at least 50% of the best-known hypervolume for the 50 MC sampling case. Prior studies have consistently shown that providing the Borg MOEA with more NFE strongly enhances its auto-adaptive features and consequently the quality of its search (Hadka and Reed, 2015, 2012b).

Transitioning to panel (b) of Figure 5.6, the computational demands increase tenfold with the 500 MC sampling scheme. Each function evaluation consists of 500 MC simulations within the same constrained 10 hour wall-clock time (i.e., fewer but less approximate function evaluations). In panel (b), the number of successful algorithmic configurations decreases significantly with the reduced NFE available for search. However, high performance is maintained for the 128 and 256 core counts exploiting 1 to 4 masters.

The importance of the cooperative, auto-adaptive search of the multi-master configurations is illustrated in panel (c) of Figure 5.6. The 1000 MC sampling scheme is the most computationally demanding variant of the LSRB test case, as successful as algorithmic performance is observed only for the L, H, and D Borg MOEA configurations using 256 cores. It is interesting to note that although configuration D exploits more NFE relative to the L and H configurations, the differences in their relative hypervolume performances are small. This again highlights the potential for the multi-master Borg MOEA to contribute purely algorithmic speedup benefits when compared to the single master-worker configuration using the 256 core counts.

The results in Figure 5.6 assume that the best Pareto approximate reference sets for each algorithm configuration are approximated across 25 random search

trials. Although often neglected in the water resources literature, the need to use multiple random search trials represents a severe computational cost, especially when exploiting parallel computing systems where users have limited allocations of computing hours. 5.6 provides a more detailed analysis of the expected effectiveness and reliability of the alternative parallel configurations of the Borg MOEA if they were run for a single random search trial by displaying the probability of attaining different hypervolume performance thresholds with each multi-master configuration. In this figure each of the columns in panels (a)-(c) correspond to an alternative parallel Borg MOEA configuration. Within each column, the white circle indicates the best overall hypervolume metric value attained by a single trial with that configuration and the color scale indicates the probability that a single trial run with that configuration attains the relative hypervolume value shown on the y-axis. Therefore, a completely dark blue bar filled to the 100% level would indicate ideal performance, designating that a single trial run of an algorithm configuration is both perfectly reliable and effective, always achieving perfect relative hypervolume. The legend in Figure 5.7 provides a key for identifying the different parallel Borg MOEA configurations tested. Again, the hypervolumes reported in panels (a)-(c) of Figure 9 are specific to a given MC sampling scheme.

The attainment results shown in panel (a) of Figure 4.4 correspond to the 50 sample MC case (i.e., the fastest but most approximate evaluations). Supporting prior observations, the fast but approximate function evaluations yield high algorithmic reliability across several parallel Borg MOEA configurations for core counts between 64 and 256. These configurations clearly have performed sufficient NFE during the 10 hour wall-clock time for the LSRB test case. It should be noted, however, that there are non-negligible hypervolume differences be-

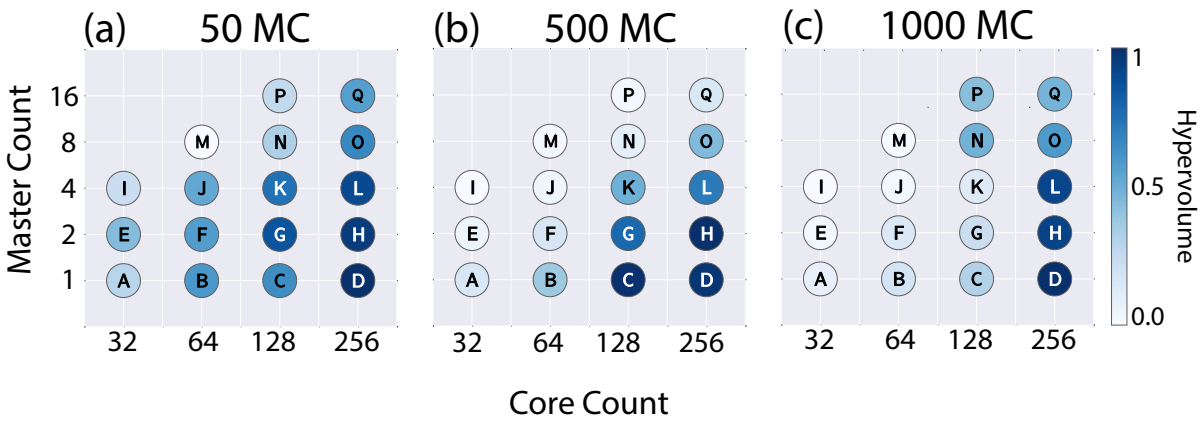

Figure 5.6: Best hypervolume (HV) performance for each combination of master and core count across the 25 random seed trials. High HV values are depicted in dark blue and low HV values are depicted light blue.

tween each configuration's best single run performances (i.e., white circles) and its reliability attainments (i.e., highest hypervolume with dark blue shading). This highlights that the single best run is often a very misleading result. For example, configuration D (master-worker with a 256 core count) attained the best overall single run for the 50 sample MC results, but the reliability of this master-worker configuration in attaining such high hypervolume performance is extremely low. Moreover, the variance in its performance is high, adding significant uncertainty in how the users would quantify the LSRB test case's tradeoffs.

Clearly the attainment plots in Figure 5.7 provide a broader context of probabilistic search performance as opposed to the reference set analysis in Figure 5.6. If a parallel configuration exhibits outstanding performance in its best single run, but fails to consistently attain high performance, it is of low value to users. For the 500 and 1000 sample MC cases shown in panels (b) and (c) of Fig-

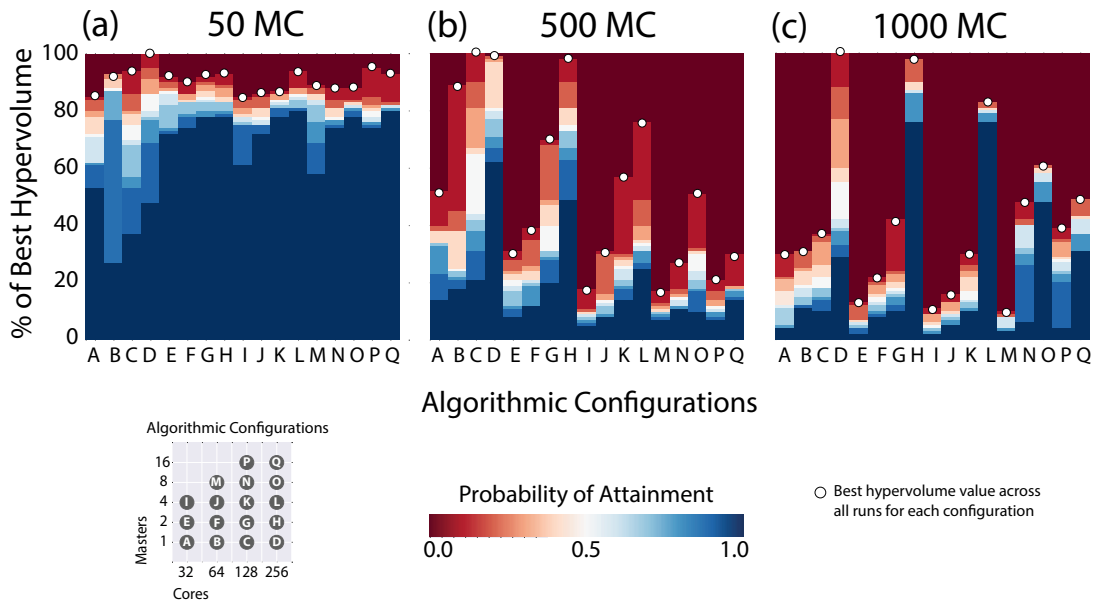

Figure 5.7: Best single run hypervolume and probability of hypervolume attainment. The white dots represent the best overall performance for the algorithmic configuration. The vertical axis shows the percent of the best hypervolume and the color scale shows the probability of attaining a given level of performance. Perfect reliability would be depicted by a blue bar filled to the 100% level with a white dot on top.

ure 4.4 respectively; their increased computational demands strongly degrade the performance of most of the tested parallel configurations of the Borg MOEA. Two significant factors shape this result. First, the obvious reduction in the NFE available for search requires dramatically more efficient search pathways for attaining high hypervolume performance. Second, the minimax formulation of the LSRB test case's objectives makes identifying control policies more challenging as more severe droughts are sampled in the 500 and 1000 sample MC cases. In panel (b) of Figure 4.4 the best run for the 500 sample MC case is obtained by configurations C (the single master-worker implementation using 128 cores). However, both the master-worker and 2-master variants of the parallel

Borg MOEA using 256 compute cores (i.e., configurations D and H in Figure 4.4) have high hypervolume attainment probabilities (or best single run reliability). As we further increase the computational demands and problem difficulty by moving to the 1000 sample MC case in panel (c) of Figure 4.4, the advantage of the auto-adaptive, interacting search of the multi-master parallelization scheme becomes more evident. Note that for the 256 core count configurations, while the master-worker (configuration D) obtained the best single trial run, the 2 and 4 multi-master configurations (H and L) had a much higher reliability in attaining improved hypervolume performance across all of their individual random trials. As noted in the results of Tang et al. (2007), master-worker parallelization schemes by themselves do not fundamentally change the search dynamics of MOEAs; they simply allow for more NFE per unit wall-clock time, which for relatively easy problems and effective algorithms may be sufficient to attain high levels of performance. However, difficult problems require fundamental changes in the underlying search operations to yield effective and reliable search. There is a broad body of literature on parallel evolutionary search (see review by Alba et al. (2013)) that has shown the island or multi-population parallelization scheme can, when employed well, improve our abilities to solve challenging problems when employed well. The cooperative, auto-adaptive search of the 2- and 4-master configurations (H and L) in panel (c) of Figure 4.4 is yielding a significant algorithmic benefit when addressing the most challenging instance of the LSRB test case.

5.5.2 Search Efficiency & Runtime Variability

Building on the attainment results in Figure 4.4, this section contributes a more detailed analysis of the runtime search dynamics. Overall, the most reliable search was attained by the 256 core implementations of the single master-worker Borg MOEA as well as the 2- and 4- multi-master configurations (i.e. D, H, and L, respectively). Panels (a)-(c) in Figure 5.8 show the runtime hypervolume dynamics for the different MC sampling cases given the same 10 hour wall-clock time. Recall that a hypervolume of 1 indicates that a given Pareto approximation has the same hypervolume as the best-known reference Pareto front for its corresponding MC sampling level. The solid lines represent the median hypervolume search progress across the random trials for each of the parallel configurations. The shaded regions bound the 10^{th} and the 90^{th} percentiles of hypervolume performance. As expected, panels (a)-(c) in Figure 5.8 show that both the degrees of approximation (i.e. the MC sampling level) as well as the parallelization configuration control the NFE that can be performed within the fixed wall-clock time. The 2 and 4-master configurations of the Borg MOEA permit more auto-adaptive, cooperation in their search but they allocate fewer cores for performing function evaluations. Consequently, the single master-worker will perform more function evaluations relative to the 2 and 4-master instances.

In panel (a) of Figure 5.8, the median hypervolume value for the master-worker configuration asymptotically approaches 90% and makes no further improvements throughout the run. The 50 sample MC master-worker configuration performs nearly 5 million function evaluations, but all search beyond 2 million evaluations is a wasted use of computational resources. As was also illus-

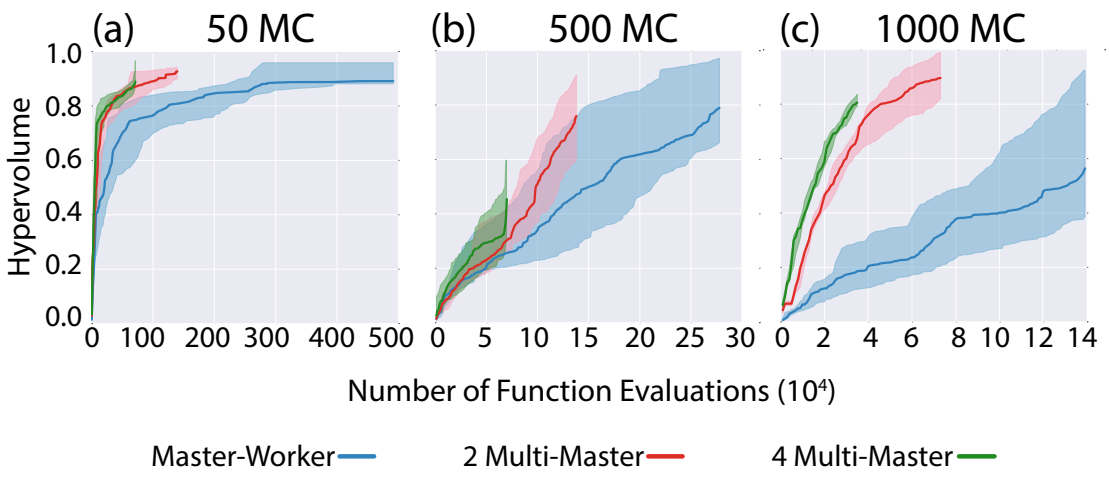

Figure 5.8: Runtime hypervolume dynamics for the master-worker, and for the 2 and 4 multi-master configurations. The horizontal axis shows the function evaluations performed by each of the problem formulations during the 10 hour wall-clock time; hence the NFE count decreases as the MC sample size increases. The vertical axis shows the runtime HV normalized against the best HV value at the end of the run for each MC sampling scheme. The solid line represents the mean across random seed trials bounded by the 5^th and 95^th percentiles.

trated in the prior work of Tang et al. (2007), the primary benefit of running the master-worker longer is reducing the variability or uncertainty of search performance across the random trials (i.e., making the results more reliable and less sensitive to random initialization). The 2- and 4 -master configurations shown in panel (a) for the 50 sample MC case rapidly achieve higher median hypervolume performance with far fewer NFE, while also significantly reducing random seed variability. These results explain why the 2- and 4-master parallel implementations had more reliable attainment results in Figure 5.8 (see configurations H and L in panel (a)).

Panels (b) and (c) in Figure 5.8 show a more pronounced improvement in

search efficiency and reliability with the multi-master cases. Given the increasing difficulty of transitioning to the 500 and 1000 sampling levels, all of the parallel configurations of the Borg MOEA are forced to exploit far fewer evaluations while seeking to quantify the LSRB test case's tradeoffs. In both the 500 and 1000 sample MC cases, the master-worker configuration has substantially less efficient hypervolume search progress throughout the run, as well as very high random seed variability. For the 500 sample MC case, the upper bound on hypervolume progress of the single master-worker configuration overlaps with the lower bound performance of the multi-master configurations. Put more simply, the best random trials of the master-worker implementation have a similar performance to the worst random trials of the multi-master configurations. Clearly, cooperative, auto-adaptive search across the masters and the controller is providing significant algorithmic search benefits that compensate for their lower NFE. The benefits of coordinated search with masters become even more pronounced for the most challenging 1000 sample MC case. In panel (c) of Figure 5.8, the master-worker configuration struggles with the higher difficulty of the 1000 sample MC case, yielding limited hypervolume progress and very high random trial variability. The 2- and 4-master configurations have dramatically lower variability throughout their run and quickly achieve high quality solutions after a substantially lower NFE count. The multi-master Borg MOEA avoids sensitivities to bad initial conditions and stalled search by effectively sharing global knowledge via the controller (see Figure 5.3). As has been shown in prior studies, the multi-master Borg MOEA quickly detects and corrects evolutionary search processes without wasting significant computing resources (Hadka and Reed, 2015).

As opposed to typical speedup calculations with a serial algorithm perfor-

mance baseline, here the master-worker configuration of the Borg MOEA is used as a baseline. For the speedup results shown in Figure 5.9 panels (a)-(c), the core count was fixed to 256 and the wall-clock time was constrained to 10 hours. Hence, the results from this section reflect the speedup related entirely to differences in algorithmic search dynamics enabled by the multi-master configurations. As discussed in section 5.3.4, the speedup calculations account for the NFE needed to attain improving levels of hypervolume performance. For each level of hypervolume, we quantify the ratio between total NFE required by the 2- or 4 -master configurations relative to the NFE required by the master-worker configuration. Any speedups therefore, require that the multi-master configurations attain improved hypervolume performance in less NFE than the master-worker baseline algorithm. Figure 5.9 panels (a)-(c) show the worst case hypervolume runtime performance for each algorithm configuration across the different MC sampling cases. In these panels, the dotted lines indicate the zones of hypervolume performance used for the speedup calculations. A broader performance threshold was established for the 50 MC configuration (0-0.9) because the master-worker configuration was able to attain high hypervolume performance. The hypervolume ranges explored for the 500 and 1000 sample MC cases are far more restricted: 0-0.3 and 0-0.35, respectively. A combination of computational constraints on NFE and reduced master-worker performance restricted the comparative ranges where all algorithm configurations had hypervolume performance values that could be compared. Note that the analysis focuses on the worst performing seeds because as an algorithm configuration's search reliability decreases, the worst case downside of its uncertain performance is of direct concern.

Analyzing the algorithmic speedups for the 50 and 500 sample MC cases

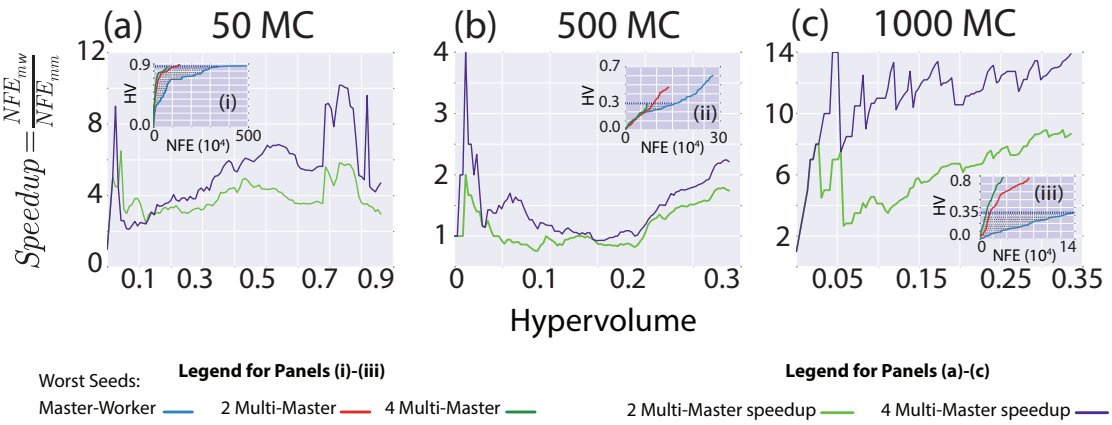

Figure 5.9: Algorithmic speedup for the 2 and 4 multi-master configurations. Panels (i)-(iii), depict the hypervolume (HV) thresholds used for the algorithmic speedup computation. Panels (a)-(c) show the ratio of master-worker to multi-master NFE performed within the specified hypervolume threshold.

in panel (a) and (b), respectively, of Figure 5.9, the 2- and 4-master configurations strongly distinguished themselves at the beginning and end of the runs. Algorithmic speedup at the beginning of the search arises from the increased diversity and early search course corrections of the multi-master configurations. Improved speedups at later stages indicate search progress for the single master-worker implementation of the Borg MOEA reaches diminishing returns far sooner than the multi-master implementations, highlighting the effectiveness of the multi-master coordination in overcoming search stagnation. Transitioning to the 1000 sample MC case shown in panel (c) of Figure 5.9, it is observed that an order of magnitude speedup using the 4-master configuration, i.e. the hypervolume achieved by the 4-master configuration requires 100,000 NFE by the master-worker configuration.

These results show that the multi-master Borg MOEA's parallel architecture

is competitive, if not superior, to the master worker configuration, even in the least NFE-constrained 50 sample MC case. As computational and mathematical difficulty increase, its search efficiency and reliability improve dramatically relative to the master worker configuration. Although these insights are in agreement with prior results on very large clusters with 10,000-500,000 compute cores (Hadka and Reed, 2015; Hadka et al., 2013), this study helps to generalize these results since the LSRB reservoir control problem is a new and different problem context relative to the problems tested in prior MOEA diagnostic studies, namely in Hadka and Reed (2015); Reed et al. (2013c). Moreover, the relatively small cluster used in this study strongly disadvantages the multi-master Borg MOEA from an NFE perspective, making it even more important for the auto-adaptive, cooperative search features to provide significant algorithmic benefits.

5.5.3 Consequences of Sampling & Algorithmic Choices

So far, we have compared the performance of algorithmic configurations within each of the MC approximation levels, but not across them. In this section, we explore the decision relevant consequences of users choosing the different levels of approximation in the MC sampling schemes in combination with the alternative parallel Borg MOEA configurations. More simply, we would like to reliably and efficiently capture the best representation of the LSRB test case's operational tradeoffs. In the context of comparing the effects of the different levels of fidelity in estimating the probabilistic evaluations of objectives, it is important too see if the more approximate 50 sample MC scheme yields appropriate representations of the LSRB test case's tradeoffs. For this assessment, we selected the reference sets across all seeds and all the algorithmic configurations for each MC approx-

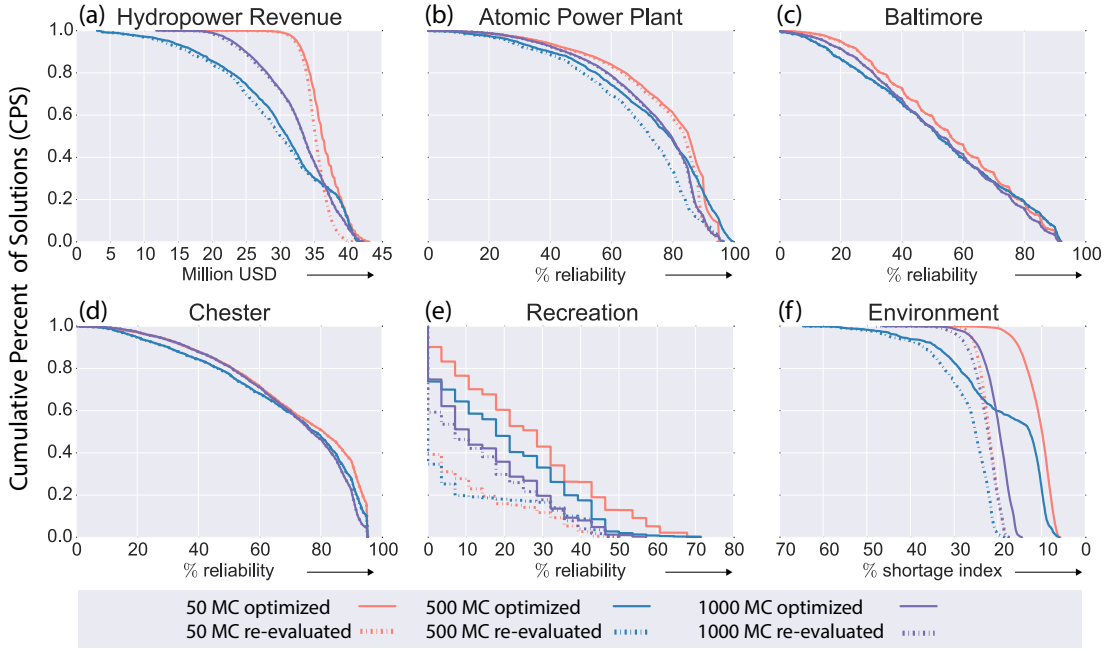

Figure 5.10: Cumulative percent of solutions (CPS) for each objective of the LSRB system. All the optimized sets (depicted by solid lines) were re-evaluated under 1000 independent MC samples (dotted lines).

imation level (i.e., the best known tradeoffs found at each MC sampling level). As discussed in prior results, the 50 and 500 sample MC configurations have the advantage of being able to perform more NFE compared to the 1000 MC formulation, but at the cost of more approximate function evaluations. These best known LSRB tradeoff solution sets for each MC approximation level were re-evaluated under an independent 1000 sample MC verification set to evaluate the consistency of representations of key tradeoffs. The results of this analysis are shown in Figure 5.10.

Panels (a)-(f) of Figure 5.10 show the cumulative percent of solutions attaining different performance levels on each of the objectives when evaluated over

the optimization and verification MC sets for the objective values over the optimization MC samples are shown with solid lines, and their values over the verification MC samples are shown with dotted lines. The arrows show the direction of preference for each objective. In panel (a) of Figure 5.10, which illustrates the relative stability of performance in hydropower revenue, we observe that the 1000 sample MC case yields re-evaluated objective values that are very similar to the optimized values. The verified hydropower performance of the reservoir control policies found using the 50 and 500 MC sample cases, however, show shifts of almost 5 million USD in magnitude for solutions with higher hydropower revenues. Relative to the maximum revenue cases, this deterioration represents a 10% loss in performance in the re-evaluated solutions.

In panel (b) of Figure 5.10, the reliability of supplying cooling water to the nuclear power plant is stable for the 1000 sample MC case while the 50 and 500 sample MC cases show degradations in performance. The solutions from the 500 sample MC case shows some degradations as high as a 10% reduction in the reliability of supplying cooling water. It should be noted that in the decision context of the LSRB, even a one percent change in reliably providing cooling water to the nuclear power plant is concerning; the Susquehanna River Basin Commission has the expressed goal of maintaining cooling water reliability for the nuclear power plant near 100%. In panels (c) and (d) of Figure 5.10, the water supply objectives for Baltimore and Chester do not shift noticeably with re-evaluation.

In panel (e) of Figure 5.10 the reliability of maintaining goal reservoir levels for the recreation objective is limited in the optimized solutions, none of which exceed 72%. Compounding the poor performance in optimization, the solutions

from the 50 and 500 sample MC cases degrade even further in re-evaluation. The optimized solutions from the 1000 sample MC case are far more stable in reliably maintaining recreation reservoir levels.

Finally, panel (f) of Figure 5.10, shows that the federally mandated environmental flows captured in the environmental shortage objective are challenging. None of the best known reference sets achieve a 0% environmental shortage even under the optimization case. The 50 and 500 sample MC cases find solutions that drastically degrade when re-evaluated, with shortages nearly doubling. The solutions found using the more challenging 1000 sample MC evaluations actually highlight the strong potential for shortages (i.e. they have encountered more challenging drought cases). The re-evaluated 1000 sample MC in panel (f) degrades slightly but is overall far more consistent.

Overall, Figure 5.10 highlights that for the control policy search context of the LSRB test case, when evaluating the computational tradeoff between more approximate but fast evaluations versus increasing the approximation level of the MC sampling, low sampling rates can yield control policies that do not consistently portray key system tradeoffs. Moreover, these results highlight the importance of verifying that the control tradeoffs presented to stakeholders are stable representations of the system tradeoffs and do not distort decision support negotiations or recommendations (see also Quinn et al. (2017b)). Figure 5.11 simulates an interactive decision support exploration of the LSRB test case's key tradeoffs where the stakeholders have expressed two performance requirements: (1) the reliability of supplying cooling water to nuclear power plant has to be greater than 95% and (2) deviations from federally mandated environmental flows have to be less than 15%. The major goal in using MOEAs

in water resources is to allow this type of interactive *a posteriori* analysis (i.e., brushing or eliminating all solutions that fail to meet these requirements). Often neglected in the literature is that our computational choices (algorithm selection, MC sampling, parallelization, etc.) can distort our perceptions of the tradeoffs themselves. We illustrate this in Figure 5.11 by creating a competition across the worst case trial runs of the top performing parallel Borg MOEA configurations such that only those configurations with re-evaluated solutions that meet the cooling water and shortages requirements are allowed to contribute to the plot. We are requiring consistent tradeoff representations that emerge for efficient, effective, and highly reliable search to attain solutions in a zone of high performance with strong objective conflicts. In short, Figure 5.11 asks, "What is the worst we can do in understanding our tradeoffs in the context of our requirements?"

In Figure 5.11 each of the vertical axes represents an objective where moving upward is always preferred. Hence, the ideal solution would be a straight line that horizontally intersects every axis at its top value. Each of the lines shown in Figure 5.11 represents a candidate solution whose re-evaluated performance meets the specified performance requirements. The requirements for the reliability of supplying cooling water to the nuclear power plant as well as the environmental flow shortages are shown brushing bars that define acceptable ranges in performance. In Figure 5.11, lines intersecting the axes below their top values are showing a reduction or compromise in those objectives. Steep diagonal lines between the axes designate strong tradeoffs. The colors of the lines indicate under which parallel configuration and re-evaluated MC sampling scheme the solutions. The colored pie chart legend shows the percent contribution of the plotted LSRB tradeoff solutions were discovered. Note that only the 2 and

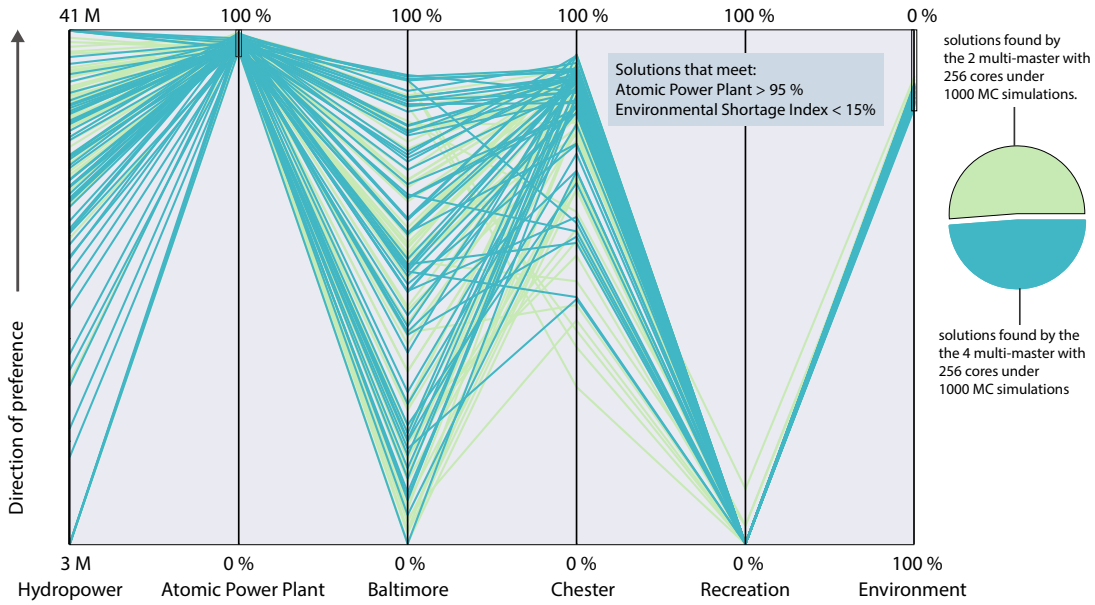

Figure 5.11: Parallel axes representation of the LSRB tradeoffs. A horizontal line along the top of the axis would depict ideal performance across all objectives. All the configurations tested in this study were re-evaluated under a common sampling scheme of 1000 MC simulations. The plot shows the solutions that yielded nuclear power plant reliability larger than 95% and Environmental shortage index less than 15%.

4 master configurations running on 256 compute cores and employing the 1000 realization sampling scheme contributed solutions to Figure 5.11. Even for their absolute worst case runs, these configurations were able to reliably find stable solutions for the most challenging case within a limited wall-clock time. Although it may seem somewhat counterintuitive, making the LSRB control problem more mathematically and computationally challenging with the 1000 MC formulation actually improved the multi-master Borg MOEA's value by triggering its auto-adaptive, cooperative search operators.

5.6 Conclusions

This study demonstrates the benefits of cooperative parallel MOEA architectures in reliably and effectively finding many objective control policies using Evolutionary Multi-Objective Direct Policy (EMODPS) Search with approximate MC evaluations. The analysis focuses on the Lower Susquehanna River Basin (LSRB) system where multiple competing objectives for hydropower production, urban water supply, recreation and environmental flows need to be balanced. We consider the tradeoffs between a better representation of uncertainty in the LSRB system and simplifying this representation to shorten the function evaluation time and allow for greater search when finding Pareto approximate control policies. We used more samples in MC simulations to increase the fidelity and robustness of evaluations of objectives, at the expense of growing computational demands and increasing the mathematical difficulty of performing well in many more extreme worlds. Counter to standard assumptions, simply increasing the number of evaluations does not necessarily improve the reliability of finding high quality and stable representations of key tradeoffs. This study demonstrated that increasing MC sampling in combination with auto-adaptive, coordinated parallel search can dramatically enhance the efficiency and reliability of search while better representing SRB tradeoffs.

Overall, the 2 and 4 island multi-master Borg MOEA solved the problem with the highest-quality results within a 10-hour wall-clock time with 256 cores. The algorithmic speedup measure also demonstrated that the previous two multi-master configurations show high benefits entirely due to cooperative search between islands. This avoids the need for running large numbers of replicate random trials; hence, we can employ less computational effort to account

for MOEAs' stochastic use of operators and random initial populations. The computational efforts can instead be employed toward improving model accuracy and better representing hydro-climatic uncertainties or evaluating other sources of uncertainties.

The algorithmic configuration also proved to have crucial impacts in the decision space, where high quality tradeoffs were observed for critical objectives, such as the Nuclear Power Plant reliability, where a percent change can have a large impact on the safety of the system high quality solutions for such critical objectives were discovered.

This study is in agreement with a long history of work (e.g., the recent review by Alba et al. (2013), also in Hadka and Reed (2015)) that shows hierarchical hybrid parallelization schemes like the multi-master schemes shown here fundamentally increase the mathematical difficulty of applications that can be addressed efficiently and reliably. This work shows that the multi-master implementations of the Borg MOEA generalize its auto-adaptivity and add cooperative search operators that allows users to reliably solve complex water systems problems. The algorithmic benefits allow to evade against not knowing how easy or difficult an application is in advance while also making it feasible to better incorporate uncertainties within the search.

The key insights of this study are useful for water managers determining how to optimize operating policies in river basin systems with multiple reservoirs and uncertain inputs impacting their operations. In these systems, EMODPS solution strategies require MOEAs that are highly scalable and extensible to emerging parallel computing architectures, and the results indicate that algorithms with multi-population architectures best meet these needs. These

algorithms will become even more useful as the availability of multiple cores in desktop, laptop, and even smartphones platforms continue to grow and increase opportunities to utilize parallel metaheuristics in water resources applications.

CHAPTER 6

ADAPTING RESERVOIR OPERATIONS TO BALANCE MULTI-SECTORAL DEMANDS DURING FLOODS AND DROUGHTS

6.1 Abstract

This study characterizes how changes in reservoir operations can be used to better balance the need for flood protection and the conflicting multi-sectorial demands in the Lower Susquehanna River Basin (LSRB), USA. Tensions in the LSRB are increasing with urban population pressures, evolving energy demands, and growing flood-based infrastructure vulnerabilities. This study explores how re-operation of the Conowingo Reservoir, located in the LSRB, can improve the balance between competing demands for hydropower production, urban water supply to Chester, PA and Baltimore, MD, cooling water supply for the Peach Bottom Nuclear Power Plant, recreation, federal environmental flow requirements and improved mitigation of growing flood hazards. The LSRB is also one of the most flood prone basins in the US, impacted by hurricanes and rain-on-snow induced flood events causing on average $100 million in economic losses and infrastructure damages to downstream settlements every year. We define a compromise policy for the Conowingo reservoir's multi-purpose operations during drought events and flood protection during wet conditions. Evolutionary Multiobjective Direct Policy Search (EMODPS) helped discover robust operating policies that were able to meet required performance for dam protection and supply reliability for Peach Bottom. They were also able to withstand pronounced drought and flood scenarios. An abstraction of the Conowingo historical operations allowed comparison against a selected EMODPS compromise

policy; were the latter outperformed historical operations when evaluated under a broader suite of hydrologic conditions. This study captures the problem of flood protection in reservoir systems with complex multi-sectorial demands, which is broadly relevant to developed river basins globally.

6.2 Introduction

Water resources managers have long confronted the challenge of seeking to improve river basin management of multi-sectorial demands and the impacts of hydrologic variability through optimized reservoir operations (Yeh, 1985; Labadie, 2004; Maass et al., 1962). Finding robust reservoir operations that are capable of withstanding hydrologic extremes is key in water resources management. This challenge is amplified by changes in the hydrologic regimes, as greater global warming increases both evaporation and atmospheres capacity to store water, leading to longer, more severe droughts and more intense flooding (Trenberth, 2011). Overcoming the negative impacts and water conflicts associated with hydrologic extremes requires innovations in current water management policies that are able to withstand severe flooding and drought conditions (Tanaka et al., 2006; Giuliani and Castelletti, 2016; World Bank, 2016). In recent years, methodological advances have emerged for the design of optimal reservoir control policies that better address operational challenges or pressures for re-operating reservoirs (Koutsoyiannis and Economou, 2003; Castelletti et al., 2012a; Giuliani et al., 2014a, 2015b). In particular, evolutionary multi-objective direct policy search (EMODPS) has emerged as a significant methodological innovation. It has shown capability of aiding in the discovery of reservoir control policies across a broader range of objectives, information sources,

and uncertainties than traditional optimization methods (Giuliani et al., 2015b; Zatarain Salazar et al., 2017). As simulation-based framework EMODPS easily exploits Monte Carlo frameworks when optimizing policies across a diversity of stochastic objectives and easily integrates exogenous variables to condition reservoir release decisions. Additionally, it eliminates the requirement of time-separable objectives, which makes it possible to simultaneously reflect different risk attitudes with respect to different system's objectives. This means that within a single problem formulation, different degrees of risk aversion can be considered for different objectives. For certain objectives, it may be more relevant to use a "risk-neutral" attitude focusing on expected values (e.g., annual revenue) whereas robustness objectives that are more "risk-averse" are better suited for extreme and lowprobability impacts. Generally, the way in which decision makers perceive risk can result in different planning strategies (Brekke et al., 2009). In this regard, EMODPS provides the flexibility to tailor risk attitudes across sectors while improving our ability to accurately assess system performance. Given that most river basins serve multiple purposes, optimizing operations results mathematically challenging solely considering the competing objectives and the high-dimensional and stochastic nature of operations. Problem difficulty can be further increased by performance constraints due to regulatory requirements or risk-aversion, severely limiting the number of feasible solutions (Characklis et al., 2006; Kasprzyk et al., 2012; Brekke et al., 2009).

Closing the gap between implementation of discovered optimal policies poses a different challenge. Reservoir operators are generally skeptical of sophisticated decision tools to base release decisions (Simonovic, 1992; Soncini-Sessa et al., 2007). Indeed, capturing reservoirs' multiple purposes and their vulnerability to uncertain conditions requires sophisticated system abstractions

to help capture their complexity. In practice, reluctance to embrace advances in robust multi-objective policy design will hinder the ability to properly manage multi-purpose reservoirs. Especially since established control rules are generally specified for single objective operations (e.g., maximizing hydropower production using limited observed hydrology). These rules fail to address emerging water needs for reservoir systems and are unable to adapt to changing hydrologic regimes. This renders vulnerable policies that may lead to performance losses, intensifying tensions across sectors under extreme hydrologic events. *Ex post* analysis has served researchers to assert the validity of improved policies which have better use of information relative to baseline operations. This style of analysis has been identified by the National Research Council (NRC) as a priority area in water resources research to test the effectiveness of water policies in the past (National Research Council, 2009). This requires abstracting historical operations. A few studies have attempted to approximate historical reservoir operations to serve as baseline when defining new control rules (Guariso et al., 1986; Giuliani et al., 2015b). These studies have demonstrated the vulnerability of *status quo* operations when tested for a larger set of objectives and hydrologic states. In this study, the intent is to move away from *ad hoc* release policies that may be suited only for specific emergency conditions. Instead, this study proposes the definition of integrated, multi-purpose reservoir management under challenging hydrologic extremes to help the robustness of the system. Additionally, it advances the discovery of policies that balance multi-sectorial demands and minimize flood risk in the Lower Susquehanna River Basin (LSRB) (Figure 2.1). The Conowingo reservoir is at the center of water management in the LSRB, it represents the last point of control between the Susquehanna River and the Chesapeake Bay. It regulates nearly 50% of the downstream flows pro-

ducing significant impact to the multi-sectorial water demands. Aside for the hydropower generation purpose, the reservoir needs to provide water to Baltimore, MD, Chester, PA, and cooling water for Peach Bottom Nuclear Power Plant while keeping target reservoir levels for recreation and federally environmental flows for fish passage during the migratory season, all of these are conflicting objectives, and are challenged during low flows specially in the summer and during fish migration, were low flows impact the reservoir's ability to meet all the required water demands (Figure 6.1).

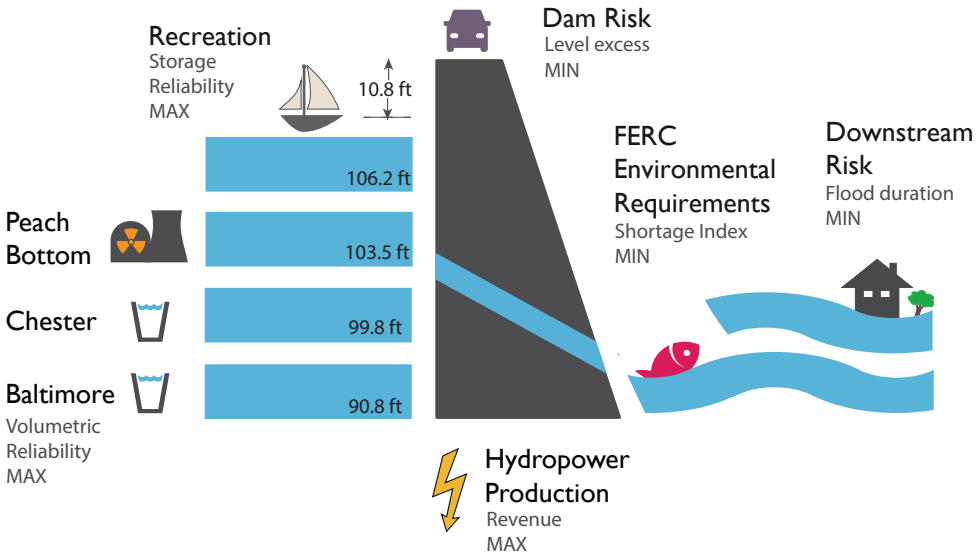

Figure 6.1: The Conowingo reservoir needs to sustain water levels to meet multiple sectorial demands for urban water supply for Baltimore, MD, Chester, PA, cooling water for Peach Bottom nuclear power plant, recreation, hydropower production and environmental flows. In addition, the Conowingo dam is the last point of control and the last barrier between the Susquehanna river and the Chesapeake bay. Operators prefer to keep water level below 109.2 ft during high flow events to prevent splashing into the road crossing above the dam. The dam is also equipped with 53 flood control gates, the gates have been opened during high flow events in the past causing damages to the immediate downstream towns.

Management in the LSRB is challenged by the complex hydrology in the re-

gion were varied topography, geology and climatic influences throw the basin into extended periods of dryness followed by pronounced storms. The climate is influenced by the Great Lakes and Midwest weather patterns in the northern-western portions and influenced by Atlantic coastal weather in the south and east portions (Neff et al, 2000). Positioning the LSRB as one of the most flood prone regions in the United States experiencing annual economic losses of $150 million every year (NOAA, 2011). Heavy rainfall and ice melt hit late winter and Mid-Atlantic hurricane season strikes in August-September (Figure 6.2). Hurricane Agnes stroke at an unlikely time, in June 1972, this was the largest recorded flood event in the region, heavy damages and loss of life motivated various structural and nonstructural solutions for flood prone areas on the main stem and the major tributaries of the Susquehanna River. Non-structural techniques included flood forecasting and warning and enhanced communication tools to offset risks associated with flooding in the SRBC (USACE, 2003; USGS, 2014). However, there are limited studies on improving operations at Conowingo to prevent downstream flooding. A report generated by Exelon (2012) studying impacts of operations in downstream flooding concluded that the most effective mitigation strategy was a structural one, stating that Conowingo has little impact in mitigating downstream flooding and instead more built storage is required. In this study, we explore non-structural alternatives by defining floodgate operations that minimize flooding severity and duration downstream.

Despite being a water rich region, the Northeast experiences droughts during peak water demands in the summer. Following a particularly dry spring, the summer of 1999 brought record-setting low flows forcing the Susquehanna River Basin Commission (Susquehanna River Basin Commission, 2006) to declare state of emergency in the river basin. The situation was extremely prob-

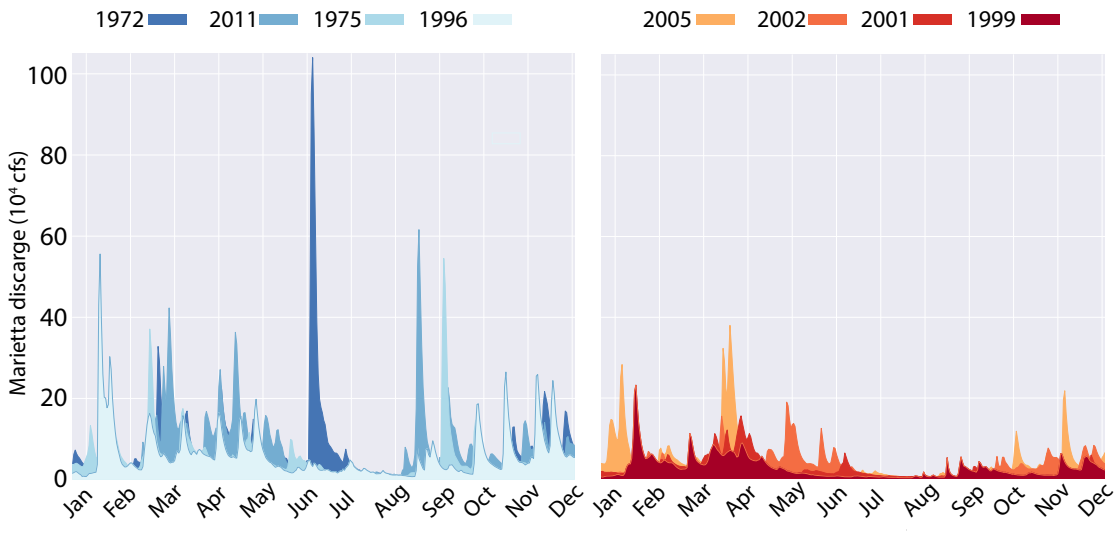

Figure 6.2: Management in the Susquehanna River Basin is challenged by extreme hydrologic conditions throughout the year. Panel (a) shows the largest streamflow events. In 1972, Hurricane Agnes hit the basin, to date is most severe flood observed in the basin corresponding to a 300-yr event. The panel also shows high inflows at the end on January during ice jams, followed by spikes in the spring during the snowmelt season. The basin is also impacted by the mid-atlantic hurricane season in September, the largest recent event was Tropical Storm Lee in 2011. Panel (b) shows the drought extremes, multiple consecutive droughts led to emergency state in the basin, with the most challenging drought observed in 1999.

lematic as the Susquehanna River serves as the primary water supply and emergency drought supply for more than 4 million consumers Susquehanna River Basin Commission (2008). The 1999 drought led to the identification of alternative management strategies (Sheer and Dehoff, 2009b). All of these alternatives were designed to manage credit for leakages and the minimum flow requirements according to the historical agreements and the regulatory constraints (Swartz, 2006). Giuliani et al. (2014b), tested the suitability of such alternatives to alleviate tensions across multiple sectors; their performance severely

degraded when tested under more severe droughts making them less effective in case of changes in the assumed hydrology. The lack of suitable policies, has generated reactive operations, were operations are only adapted for emergency conditions. This style of operations intensifies tensions and puts the LSRB at risk and generates timely and costly re-negotiations. This study advances management for the Susquehanna River Basin for sectors impacted by droughts and floods. An illustrative compromise release policy for the Conowingo system is shown to be capable of robustly balancing water demands while minimizing flooding damages to vulnerable towns downstream. The EMODPS framework facilitated the discovery of multi-purpose operating policies for Lower Susquehanna River Basin, using risk-averse measures under a broad suite of hydroclimatic conditions to better capture impacts of extremes in the system.

6.3 Methods

This study builds off the Conowingo management model described in Chapter 2 where the mass balance between the Conowingo reservoir and the Muddy Run pumped storage facility is simulated in a 4 hourly time-step to resolve short time scale floods and energy price dynamics. The Evolutionary Multi-objective Direct Policy Search (EMODPS) framework, described in Chapter 3.2.1, is used to define reservoir operations that balance multi-sectoral water demands while minimizing flood risks. In prior chapters, optimal operations for the Conowingo reservoir focused on droughts; however, floods represent a significant concern for the Lower Susquehanna River Basin (LSRB). This study extends the optimization problem formulation to also account for flood risks while balancing the prior established multi-sectorial demands. Section 6.3.1

provides a description of the objective formulations. Section 6.3.2 describes the formulation of the operating policies. Section 6.3.3 provides a summary of the synthetic hydrology generator used to better resolve extreme events (i.e., floods and droughts). Finally, Section 6.3.4 describes the approximation of historical preferences to benchmark the performance of the new control policies discovered in this work.

6.3.1 Formulation of Objectives

In this chapter, the LSRB objectives described in Chapter 2 were re-formulated to better capture a more stable and generalizable representation of the system's risk-averse tradeoffs. Quinn et al. (2017a) found that minimizing the worst first percentile across an ensemble of streamflows guarantees improved control policies that are stable and convergent relative to more traditional worst-case or min-max measures that are often used in robust optimization. Using the notation of Quinn et al. (2017a) the general formulation is represented in Equation 6.1.

$$J_d = \psi_{i\in(1,\ldots,N)}\{\Phi_{t\in(1,\ldots,T)}(g_{d(t,i)})\} \tag{6.1}$$

In this study, $g_{d(t,i)}$ is the daily average over the 4-hourly time-step for the $d-th$ objective for the $i-th$ ensemble member, Φ is the aggregation operator for selecting statistical measures of $g_{d(t,i)}$ with T=365 to better capture the inter-annual variability, Ψ is the aggregation operator across N ensembles. Minimizing the worst first percentile helps to guarantee the robustness of the LSRB control policies for a large suite of hydrologic conditions.

Maximize Hydropower Revenue

Given that the Conowingo reservoir and Muddy Run are privately owned power generators, hydropower revenue is a significant preference for the LSRB system. Hydropower revenue J_{hydro} is optimized to maximize the worst first percentile ensemble member for the total annual hydropower revenue in Equation 6.2.

$$J_{hydro} = quantile_{i\in(1,\ldots,N)}\left\{\sum_{t=1}^{T} HP_{t,i}\cdot\rho_t, 0.01\right\} \tag{6.2}$$

Where the average revenue is a function of daily hydropower production $HP_{t,i}$ for the $i-th$ ensemble member and the hourly energy prices ρ_t using the Pennsylvania-New Jersey-Maryland (PJM) energy market.

Maximize Water Supply Reliability

Our reliability objectives are motivated by Hashimoto et al. (1982b) to better capture the system's performance during drought, peak demands or extreme weather. The authors suggest that evaluating system failures in terms of their reliability, resilience and severity provide a better measure of policy performance under a wide range of future conditions. These objectives have since been widely adopted in the water resources literature, primarily with respect to failures in meeting demand. The Conowingo reservoir supplies a fraction of the urban water required for the city of Chester, PA, and for Baltimore, MD, it also provides cooling water to the Peach Bottom Nuclear Power Plant. In this study, the goal is to maximize volumetric reliability (J_{VR}) for each of the system's water

demands (d). The worst-case objective is obtained for each ensemble member and the worst first percentile is minimized with T= 365 and N=1000 in Equation 6.3.

$$J_{VR} = quantile_{i\in(1,\ldots,N)}\left\{min_{t\in(1,\ldots,T)}\left(\frac{Y^d_{t,i}}{D^d_{t,i}}\right), 0.01\right\} \tag{6.3}$$

Where $Y^d_{t,i}$ is the delivery and $D^d_{t,i}$ is the water demand for day t, for the i-th ensemble member for the d-th supply objective.

Maximize Storage Reliability

The reservoir provides an essential recreational resource for the region receiving over 250,000 visitors each year who enjoy fishing, boating and kayaking. The busiest influx of visitors is during the weekends between Memorial Day and Labor Day. During this time, the target reservoir level is kept above 106.5 ft. The structure of the recreation objective reflects its time sensitivity, capturing good performance only when the target level is met during the weekends of the touristic season. This objective also reflects the reservoirs state during critical water demand periods, namely when the recreational season overlaps with the fish migratory season. During this time the reservoir also suffers from low rainfall and high evaporation losses impacting the reservoir's storage reliability J_{SR}. In Equation 6.4, reliability is measured as the number of weekends during the touristic season (W) when the reservoir level is below the target level (TL). Again, the worst first percentile is minimized across N ensemble members for t days in the touristic weekends for the i-th ensemble member.

$$J_{SR} = quantile_{i\in(1,\ldots,N)}\left\{\left(1 - \frac{\sum_{t=1}^{13} TL}{\sum_{t=1}^{13} 2W}\right)_i, 0.01\right\} \tag{6.4}$$

Minimize Environmental Shortage Index

Reservoir operations by Exelon Corporation are subject to licensing from the Federal Energy Regulatory Commission (FERC). This license is subject to maintaining water levels for recreation and minimum flow releases as mandated by the Susquehanna River Basin Commission (2006). Current minimum flows were established to provide protection for fishery resources, with the highest flows required during the fish migratory period in the spring permitting intermittent flows solely during the winter when fish populations are limited. To capture this requirement, the squared deviations between FERC seasonal requirements $Z_{t,i}$ and the actual daily releases $Y_{t,i}$ are measured. Were the maximum deviation across t days with T=365 for the i-th ensemble member. The worst first percentile environmental shortage index J_{SI} is minimized across N ensemble members to ensure that 99% of the shortages are at least as small when optimizing operations.

$$J_{SI} = quantile_{i\in(1,\ldots,N)}\left\{max_{t\in(1,\ldots,T)}\left(\frac{max(Z_t - Y_t, 0)}{Z_t}\right)_i^2, 0.99\right\} \tag{6.5}$$

Minimize Level Excess

Operations at the Conowingo reservoir need to balance the previous multisectorial water demands while managing storage levels to accommodate high inflows during extreme rainfall and snowmelt events. Reservoir operators at-

tempt to keep water surface elevation at a goal height 109.2 ft out of the maximum structure height of 120 ft available. The goal reservoir level provides a safety board and prevents splashing into road US 1 crossing on top of the dam. To reflect dam safety, deviations are minimized from the level at Conowingo (h_t^{CO}) when it exceeds the desired level (h^{crit}) in Equation 6.6. Similar to the shortage index, the maximum annual deviation is obtained across t days with T=365, for the i-th ensemble member; subsequently, the worst first percentile level excess (J_{LE}) is measured in Equation 6.7.

$$\Phi^{LE} = \begin{cases} h_{t,i}^{CO} - h^{crit}, & \text{if } h_{t,i}^{CO} > h^{crit} \\ 0, & \text{otherwise} \end{cases} \tag{6.6}$$

$$J_{LE} = quantile_{i\in(1,\ldots,N)}\{max_{t\in(1,\ldots,T)}(\Phi^{LE}(i)), 0.99\} \tag{6.7}$$

Minimize Flood Duration

The Conowingo reservoir represents the last flow control barrier between the LSRB and the Chesapeake Bay. The reservoir is equipped with 53 flood control gates to manage releases from the Susquehanna River during high flow events. As the number of these gate openings increases there are direct flooding threats to the town of Port Deposit, located only 5 miles downstream from the dam. This town has had a history of flooding since the 1800s, it suffered flooding nearly every spring and when the river thawed. The construction of the dam has helped regulate the downstream flows preventing the frequency in which such flooding happens; however, some of the most damaging floods were registered after its construction. Based on historical records, Port Deposit generated

a description of the damages expected downstream depending on the number of gates opened at Conowingo (Figure 6.3).

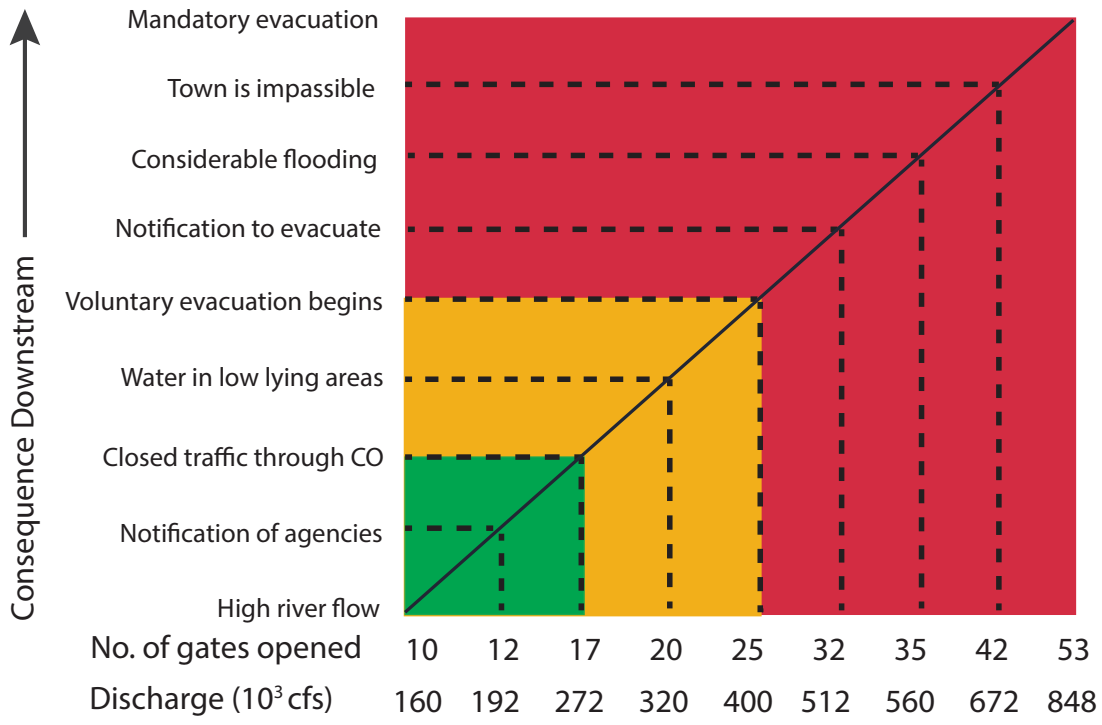

Figure 6.3: The town of Port Deposit located at only 5 miles downstream from the Conowingo reservoir receives direct impacts from gate openings at the dam. The chart shows the description of the consequences to the town relative to the number of gates opened at Conowingo. The consequences range from small in green, to moderate in yellow, to severe in red. The goal is to stay below 25 opened gates to minimize the impacts. Note that with 35 opened gates there is considerable flooding, and above 42 gates the town is impassible.

Based on this description, one prevents the occurrence of higher impacts (in red) and attempt to keep the consequences from low (in green) to moderate (in yellow) when gate openings are required. The maximum number of gate

openings is kept at 20 and the duration is minimized when this spill level is exceeded. Flood duration (Φ^{FD}) in Equation 6.8 is described by $a_{t,i}$, in Equation 6.9 which indicates if the release $r_{t,i}^{spill}$ is above the critical spill r^{crit}, and $b_{t,i}$ in Equation 6.10 indicates if the critical spill has just begun; for day t and ensemble member i.

$$\Phi^{FD}(i) = \frac{\sum_{t=1}^{T} a_{t,i}}{\sum_{t=1}^{T} b_{t,i}} \tag{6.8}$$

$$a_{t,i} = \begin{cases} 0, & \text{if } r_{t,i}^{spill} \leq r^{crit} \\ 1, & \text{if } r_{t,i}^{spill} \geq r^{crit} \end{cases} \tag{6.9}$$

$$b_{t,i} = \begin{cases} 0, & \text{if } t = 1 \text{and } r_{t,i}^{spill} > r^{crit} \\ & \text{or } t > 1, r_{t,i}^{spill} > r^{crit} \text{ and } r_{t-1,i}^{spill} > r^{crit} \\ 1, & \text{otherwise} \end{cases} \tag{6.10}$$

The 99^{th} percentile value of flood duration J_D across the N ensemble members is measured in Equation 6.11.

$$J_{FD} = quantile_{i\in(1,\ldots,N)}\{max_{t\in(1,\ldots,T)}(\Phi^{FD}(i)), 0.99\} \tag{6.11}$$

6.3.2 Defining Control Policies

Candidate LSRB control policies utilize the current time step's reservoir level and the seasonal time index as inputs for Gaussian radial basis functions (RBF)

to recommend releases. The RBF policies provide the prescribed releases u_t at time t, normalized on $[0, 1]$. These policies are de-normalized into actual water releases preserving the reservoir's physical restrictions and available water.

$$u_t = \sum_{i=1}^{A} w_i \exp\left(- \sum_{j=1}^{B-2} \frac{((x_{t-1})_j - c_{j,i})^2}{b_{j,i}^2} + (x_{t-1})_{B-1}^2 + (x_{t-1})_B^2) \right) \tag{6.12}$$

In Equation 6.12 $(x_{t-1})_{B-1} = sin\left(\frac{2\pi t}{365} - \rho_1\right)$ and $(x_{t-1})_B = cos\left(\frac{2\pi t}{365} - \rho_2\right)$. In this problem formulation, a cyclic representation of time is provided with phase-shifted $sin(\cdot)$ and $cos(\cdot)$ functions of time to guarantee continuity of annual policies when transitioning from December to January across multiple simulated years (Quinn et al., 2017a). B is the number of inputs, A is the number of RBFs, w_i is the weight of the i-th RBF, and $c_{j,i}$ and $b_{j,i}$ are the centers and radii of the j-th input for the i-th RBF. The vector of decision variables (θ) is formed by the centers, radii and weights of the RBFs and the sine and cosine phase shifts, formally defined as follows:

$$\theta = \begin{bmatrix} c_{i,j} \\ b_{i,j} \\ w_i \\ \rho_1 \\ \rho_2 \end{bmatrix} \tag{6.13}$$

with $i = \{1, ..., A\}, j = 1, ..., B$ and $k = [1, ..., M]$, where $c_{i,j} \in [-1, 1], b_{i,j} \in [-1, 1], w_i \in [0, 1]$ with $\sum w_i = 1$ and phase shifts p_1 and $p_1 \in [0, 2\pi]$. In our problem formulation $B = 3$ inputs: the storage at Conowingo and time, with phase-shifted sine and cosine functions as described above. As a general rule,

the number of the number of RBFs is $A = B + 2 = 5$, to approximate our policy and $K = 4$ releases, one release for each of the water supply demands (for Baltimore, Chester, and Peach Bottom) and the downstream releases. Downstream releases are disaggregated into turbine releases for hydropower production, minimum flows for fish passage, and spillway releases to account for flooding impacts downstream. The formulation presented here has a total of $A(K + 2(B - 2)) + 2 = 5(4 + 2(3 - 2) = 32$ decision variables. The non-dominated parameters for the candidate control policies θ^* in Equation 6.14, are identified by minimizing the eight objectives in Equation 6.15, note that all the maximization objectives are negated since the MOEA defaults to assume minimization problems.

$$\theta^* = argmin_\theta J(\theta) \tag{6.14}$$

where:

$$y = \begin{bmatrix} -J_{Hydrower}(\theta) \\ -J_{Power\ Plant}(\theta) \\ -J_{Baltimore}(\theta) \\ -J_{Chester}(\theta) \\ -J_{Recreation}(\theta) \\ J_{Environment}(\theta) \\ J_{Level\ Excess}(\theta) \\ J_{Flood\ Duration}(\theta) \end{bmatrix} \tag{6.15}$$

Subject to:

$$J_{Power\ Plant} \leq 90 \tag{6.16}$$

and

$$J_{Level\ Excess} \leq 8.8ft \tag{6.17}$$

The optimization problem is constrained as shown in Equations 6.16 and 6.17 to guarantee 1) minimum performance for nuclear power plant supply reliability and 2) to assure a safety board of at least 2.2 ft to prevent overtopping the dam. This formulation was solved using the multi-master Borg MOEA, described in Section 5.3.3, to find Pareto approximate policies relative to the system's objectives. Once finding the Pareto approximate control policies, their generalizability is assessed by simulating them with an out of sample hydrologic ensemble that is 100 times larger than the optimization ensemble. Similar to the optimization, the worst first percentile objective is computed to obtain a conservative measure of performance. If the objective values in the re-evaluation are similar to those obtained in the optimization, then both the objectives and policies are stable, so the representation of policy performance can be trusted. This signifies that the policies perform well not only with the hydrology in which they were optimized, but they sustain performance under a broader suite of hydrologic scenarios with more pronounced droughts and floods embedded in the simulation.

6.3.3 Capturing Hydrologic Uncertainty

In this study, operating policies for the LSRB's Conowingo Dam are evaluated over synthetic hydrology to better capture the system's flood and drought extremes since these conditions are rarely observed in the historical record. The synthetic generation method described in Section 3.3 was used to generate synthetic hydrology for the system, except a few updates were implemented for the disaggregation procedure. Because the proportional rescaling by month described in Section 3.3 results in scaled versions of historical flow patterns occur-

ring on the same days each year, a wider network of neighbors is now considered by including monthly totals within a moving window of +/- 7 days of the current month when selecting neighbors. That is, rather than only considering historical January flows as neighbors to the synthetic January flows, for example, total flows over 31 consecutive days within the period from the last week of December to the first week of February are considered.

Data for the generation were obtained from the USGS Marietta gauging station (01576000) and OASIS model output. Synthetic streamflows were generated for the Marietta gauging station, inflows to Muddy Run, and lateral inflows between Marietta and the Conowingo Dam, as well as evaporation rates over the Conowingo and Muddy Run dams.

Log-space historical and synthetic probability of exceedance curves for Marietta inflows in Figure 6.4 indicate that the synthetic hydrologic variables generate more extreme high and low values. Note that Hurricane Agnes, a 300-year event, is captured in our historical record, consequently, our synthetic traces generate events of this magnitude and larger enabling to define robust policies for pronounced flood events. The probability of exceedance curves show the 1000 years of synthetic hydrologic used for the optimization and the 100,000 years of synthetic hydrologies used to test the generalizability of the optimized policies.

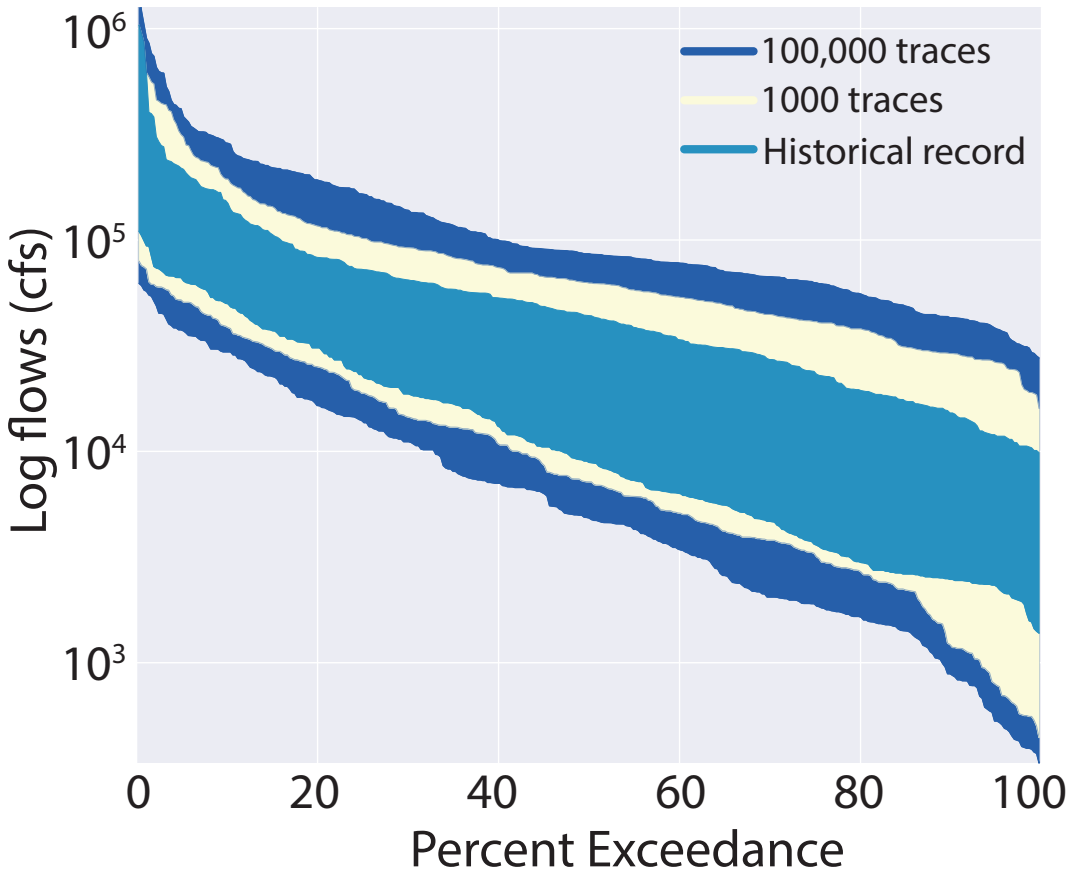

Figure 6.4: Flow duration curve ranges spanned by historical flows measured at the Marietta gaging station (in teal). Synthetically generated flows used for robust optimization (in yellow) and those used for re-evaluation of the optimized policies (in dark blue). Here, the log flows with 0% percent exceedance probability represent floods while those at 100% exceedance represent the droughts. The synthetic generation expands the extremes observed in the historical record.

6.3.4 Capturing Historical Preferences

As part of this study, an *ex post* analysis of historical operations is used to clarify the inherent value of the EMODPS compromise policy across the LSRB's objectives. The *ex post* comparative assessment focuses on challenging historical flooding and drought events. In terms of flooding, Tropical Storm Lee (September 2011), serves as our benchmark for high flow events. It represents a recent large-scale flooding event in the river basin that caused extensive flash flooding in the LSRB displacing over 100,000 people and causing an estimated of 1 billion USD in damages. Flooding was exacerbated due to consecutive events, first Hurricane Irene hit the basin in August followed by Tropical Storm Lee in September, 2011. This decreased the Conowingo reservoir's storage capacity to accommodate large inflows and led to large releases through flood control gates downstream in a short period of time.

As a challenging drought case, the historical year 1999 represents a relatively recent case where multiple consecutive dry years lead to a declared state-of-emergency in the LSRB. The Conowingo reservoir struggled to maintain FERC mandated minimum flow releases due to significant flow deficits during this time. The drought disrupted power production, limited water supply withdrawals, and strongly diminished recreational activities in the LSRB. The FERC allowed temporary waivers to include gate leakage towards meeting minimum flow releases since critical and serious impacts were developing with no projected improvement (Susquehanna River Basin Commission, 2006). Historical operational preferences were identified by fitting the EMODPS policy parameters that maximize the Nash-Sutcliffe efficiency (NSE) as proposed by Nash and Sutcliffe (1970) relative to historical downstream gauged flows below the

Conowingo reservoir. NSE is a standard metric used to assess goodness of fit in hydrologic modeling. The metric is a normalized statistic that determines the relative magnitude of the residual variance (noise) compared to the measured variance (information). The range of NSE lies between -∞ and 1, where NSE of 1 would indicate a perfect fit, whereas an NSE lower than zero indicates that the mean value of the observed time series would have been a better predictor than the model, indicating unacceptable performance. Using inflows from the Marietta gaging station located upstream the reservoir for the period of 1968-2016 to simulate the downstream releases. This policy is approximated to the gaged releases for the corresponding period measured at the USGS 01578310 gage located downstream of the Conowingo reservoir. The identified historical policy is based on a conservative evaluation that focuses on minimizing the worst-case annual NSE across the 49 years of historical record used in Equations 6.18 and 6.19.

$$J_{NSE} = min_{i\in(1,\ldots,49)}\left\{1 - \frac{\sum_{t=1}^{365}(O_{t,i} - R_{t,i})^2}{\sum_{t=1}^{365}(O_{t,i} - \bar{O}_i)^2}\right\} \tag{6.18}$$

Where $O_{t,i}$ and $R_{t,i}$ is the observed and simulated release at day t for the i-th year, and $\bar{O}_i$ is the mean for the i-th observed year. The NSE for the worst case fit is maximized using the Borg MOEA to find the best historical approximation θ^{hist} defined in Equation 6.19.

$$\theta^{hist} = argmin_{\theta}(-J_{NSE}(\theta)) \tag{6.19}$$

Similar to the optimization in Section 6.3.2, the parameters of the approximated historical policy are the centers, radii and weights for the RBF and the sine and cosine phase shifts. Again, the NSE objective (J_{NSE}) is negated for its

maximization; where the multi-master Borg MOEA was used to find the parameters that maximized the NSE values. Given the identified historical *ex post* policy's parameters, the RBF can then be used to map them into the EMODPS objective space to measure its performance relative to the optimized policies using a broader suite of stochastic hydrologic traces.

6.3.5 Computational Experiment

The multi-master Borg MOEA (Hadka and Reed, 2015), described in Chapter 5.3.3, was used to find optimal control policies for the Conowingo reservoir. The multi-master Borg MOEA is able to find solutions for this constrained, multi-dimensional problem due to its cooperative nature and adaptive use of operators. In our previous study, the multi-master Borg showed high reliability and small random seed variability across independent runs, implying that coordinated search is less dependent upon initial populations and operator probabilities. These findings enabled us to gear computational efforts towards larger exploration. In this study, a large number of function evaluations (NFE) are required to find feasible solutions that meet the two performance constraints for nuclear power plant supply reliability and dam freeboard. The multi-master Borg proven reliability enables us to reduce the number of random seed computations when accounting for variability in initial populations and stochastic search operators, and focus instead on a close representation of the complexity of our system. The problem is optimized using XSEDE Comet cluster (https://www.comet.sdsc.xsede.org) with 1728 processing cores for 48 hours. The multi-Master Borg is used, with 8 islands and 500,000 function evaluations per island. The Pareto sets consist on combining the best solution found across

the 8 islands. The epsilon dominance archiving used in Borg requires that users specify levels of precision for each objective below. Table 6.1 shows the values of epsilon (or significant precisions) used for each objective in this study.

Table 6.1: Epsilon values for each objective

Objectives	Epsilon values
$J_{Hydrower}$	0.5
$J_{Power\ Plant}$	0.05
$J_{Baltimore}$	0.05
$J_{Chester}$	0.05
$J_{Recreation}$	0.001
$J_{Environment}$	0.001
$J_{Level\ Excess}$	0.001
$J_{Flood\ Duration}$	0.001

6.4 Results and Discussion

This section presents the results from the optimization of EMODPS policies for the LSRB system and their subsequent re-evaluation. Section 6.4.1, discusses the tradeoffs that emerge from the Pareto-approximate solutions discovered under our risk-averse robust optimization formulation that are then assessed for their generality by stressing these solutions on a substantially larger sample of streamflows. Using an illustrative compromise policy that balances the Conowingo reservoirs water demands and flood risks Section 6.4.2 reveals expected annual performance for the compromise policy (i.e., a more risk neutral

assessment). Section 6.4.3 uses detailed stochastic diagnostics to illustrate how the expected dynamic behaviors of the candidate Conowingo reservoir control policies change for different performance preferences. Section 6.4.5 presents an *ex-post* assessment of how the LSRB's historical control preferences compare to discovered our compromise control policy.

6.4.1 Capturing Multi-sectorial Tradeoffs and Flood Risk for the Conowingo System

Figure 6.5 presents a parallel axis plot visualization (Inselberg, 2009) of the Pareto approximate tradeoffs between the Conowingo reservoir's multi-sector demands and flood risks. In the plot each line corresponds to an operating policy that intersects each vertical axis at the value it achieves for that objective. The vertical axes are oriented such that the optimal direction is upward; the top axis shows the smallest values achieved for minimization objectives and the highest values achieved for maximization objectives. All lines are shaded according to their performance on hydropower revenue, with darker color representing greater revenue. Consequently, a dark horizontal line intersecting the top of each axis would represent the ideal solution.

The performance tradeoffs illustrated in Figure 6.5 are the result from optimizing the worst first percentile objectives across a 1000-member streamflow ensemble and then re-evaluating the optimized policies across an independent 100,000-member ensemble. This formulation provides a highly risk averse representation of the tradeoffs, thus the performance level illustrated in Figure 6.5 are quite conservative because the operating policies achieved equivalent or bet-

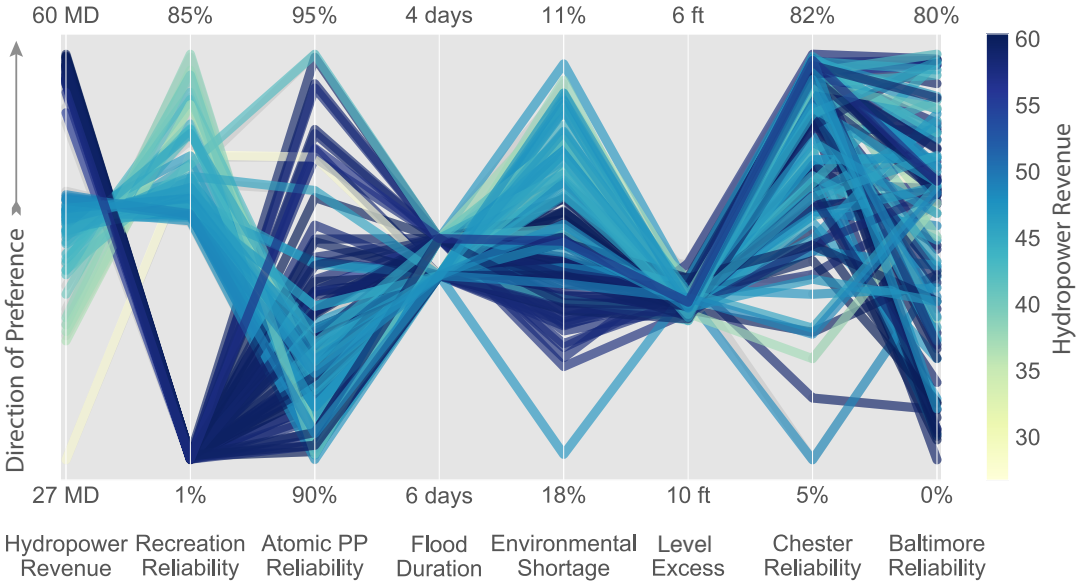

Figure 6.5: Tradeoffs discovered for the Conowingo system. Each line in the parallel axis plot represents a different policy obtained with our risk averse EMODPS formulation. The values show the result from the worst-first percentile re-evaluation across 100,000 synthetic hydrologies of our optimized policies. The direction of preference is upwards, with the best solution attained for each objective reflected in the top axis. The lines are colored by hydropower revenue; policies colored dark blue indicate high hydropower revenue.

ter performance than the depicted values for 99% of the hydrologic scenarios used in their re-evaluation. Furthermore, the re-evaluated policies met the performance criteria that guarantee supply reliability for Peach Bottom Nuclear Power Plant above 90% and level excess below 8.8 ft. As a reference, a level excess above 11 ft would indicate dam overtop. Hence, meeting the level excess constraint is a crucial safe freeboard for the dam. In Figure 6.5, all the solutions achieved at least 4.2 ft of freeboard. Additionally, the tradeoffs for the system remained nearly unchanged when re-evaluating the policies under a larger hydrologic sample than that used in their optimization, verifying that the mod-

est streamflow sampling size used in the multi-objective robust optimization was sufficient. This is one of the benefits of our worst first percentile formulation, which provides a stable and bounded objective performance (Quinn et al., 2017a). Moreover, our policies met the constraints when evaluated out of sample. This is noteworthy because our constraints are not explicitly stated in the re-evaluation step; instead, an unconstrained version of the system simulation is used to map prescribed policies into release decisions as a function of seasonality and storage. The tradeoffs in Figure 6.5 highlight that worst first percentile formulation yielded narrow performance ranges for our flooding objectives, with level excess values between 8 and 8.6 ft, and flood duration between 4.8 days and 5.10 days (4 days and 20 hours and 5 days and 2 hours). Recalling that for the re-evaluation step, the policies are tested under the synthetic hydrology shown in Figure 6.4. This means that the worst first percentile measure of the flooding objectives yields an event comparable to Hurricane Agnes in 1972. To provide context, this was a 300-yr event and is the largest observed flood in the Susquehanna River Basin with a discharge of 1.04 million cfs measured upstream the Conowingo reservoir. An event of this magnitude would naturally challenge storage limitations in the reservoir. Our policies; however, performed well enough to meet the required freeboard. As mentioned earlier, all policies attained a safety board above 2.2 ft (or a level excess $<$ 8.8ft). However, none of our policies achieved the desired reservoir level of 109.2 ft (equivalent to a level excess of 0 ft). In this study the flooding formulation used the guidelines generated by the town of Port Deposit (Figure 6.3) to measure the flood duration objective. According to these guidelines, severe flooding happens in the town when more than 25 gates open at the Conowingo reservior. The goal is to minimize the duration of discharge above 320,000 cfs, which corresponds to

20 gates opened at maximum capacity. Severe downstream flooding is avoided by penalizing performance when flood gate discharges are above the critical release and by minimizing the time when the discharge is above this level. Our policies attain a discharge between 320,000-432,000 cfs equivalent to opening 20-27 gates, this would be distributed through a period of 4 to 5 days; avoiding the damages associated to opening all the floodgates during a brief period of time. In fact, 53 gates were consecutively opened during Agnes causing severe damages downstream; this outcome is avoided with our problem formulation. Similar to the level excess objective, flood duration is measured for an event of similar magnitude to Agnes. Both of the flooding objectives have effects on each other. Namely, the level excess constraint can potentially impact flood duration since a lengthier release through floodgates would be necessary to free storage to preserve the required water level at the dam during high inflows. Intuitively, flooding focused performance constraints yield performance tradeoffs across other sectors. This is evident in Figure 6.5 for urban water supply reliability; the policies achieve variable reliability ranging from 0-80% for Baltimore, MD and 5-82% for Chester, PA. There is a broad range in performance for the recreation objective as well. This can be explained by its time sensitivity. This objective requires meeting a specific water level during a small time window (i.e., weekends of the recreational season). While the current problem formulation does not favor recreation, it remains a crucial objective since the FERC license granted to Exelon is contingent upon satisfying water elevation for recreation in the reservoir. As expected, sustaining high hydropower generation decreases storage reliability for recreational activities in the reservoir. Figure 6.5 shows a steep compromise between recreation reliability and hydropower revenue. The resulting tradeoff has been confirmed in previous formulations of the

Conowingo system (Giuliani et al., 2014a; Zatarain Salazar et al., 2017). Another important aspect of these results, is that the policies achieved supply reliability above 90% for the Peach Bottom nuclear power plant. In practice, this objective is treated as a constraint with supply reliability close to 100% required all year round. In the current study's problem formulation, nuclear power plant reliability is modeled as an objective with minimum performance goals as opposed to treating it as a hard constraint to understand how cooling water demands for Peach Bottom influence the robust multi-objective optimization's tradeoffs. Capturing these tradeoffs helps analyze the impacts for conflicting water demands that are affected by low flow conditions (e.g. hydropower production, downstream releases for environmental flows and water supply). Contrary to our flooding objectives, a worst first percentile aggregation for "drought objectives" or objectives that are impacted by drought, means that the evaluation of our policies is now performed under drought conditions that are more severe than those observed in the historical record. The streamflow used for the re-evaluation of these objectives is shown in Figure 6.4 in blue, the 100% probability of exceedance in the x-axis shows the drought scenarios. Observe that the synthetic streamflow used in the re-evaluation is significantly lower than that of history. This means that the policies were evaluated under unprecedented low flow conditions that expand beyond the historical record. This record already includes serious droughts such as that of year 1999 when the basin was declared in state of emergency due to dangerously low flows. During this time, challenging tensions emerged across the multi-sectorial demands. The severity of the conditions reduced the capability to meet the FERC requirements forcing a waiver to allow counting leakage towards the minimum downstream releases, also urban water supply was restricted. Keeping in mind the severity of the

droughts used for our re-evaluation, the highest reliability attained for the nuclear power plant objective was 95%. In practice, this would be an alarmingly low reliability; nonetheless, it helps us understand the system's capability to meet this crucial demand when it is subject to critically low flows. This outcome could be further hindered given the plausible scenario of increasing demands. Section 6.4.2 illustrates an example compromise for balancing the Conowingo reservoir's multi-sectorial demands while providing sensible flood protection.

6.4.2 Selecting a Compromise Policy

The tradeoffs illustrated in Figure 6.5 provide rich information context that can be used to inform negotiated compromise. Critical goals in this case are those that would normally be considered as constraints but are instead treated as objectives. As discussed earlier, the purpose of treating them as objectives is to visualize and understand the full tradeoffs for the system. The illustrated compromise solution therefore needs to yield high performance for the nuclear power plant, the environmental flows and level excess. Beginning with the Nuclear Power Plant, our illustrated brushed performance requirements in Figure 6.6 specify supply reliability above 94%. Subsequently, the brushed performance requirement for the environmental shortage index is below 14%. This index indicates deviations between the required flows stated by the SRBC and the actual flows sent downstream. Ideally, the flows would stay as close as possible to the requirements since large deviations can negatively affect the health of the Chesapeake Bay and impact fishing resources. Operations at Conowingo prioritize meeting mandatory environmental flows to support relicensing efforts for Exelon, which is why large deviations are unlikely to happen in practice.

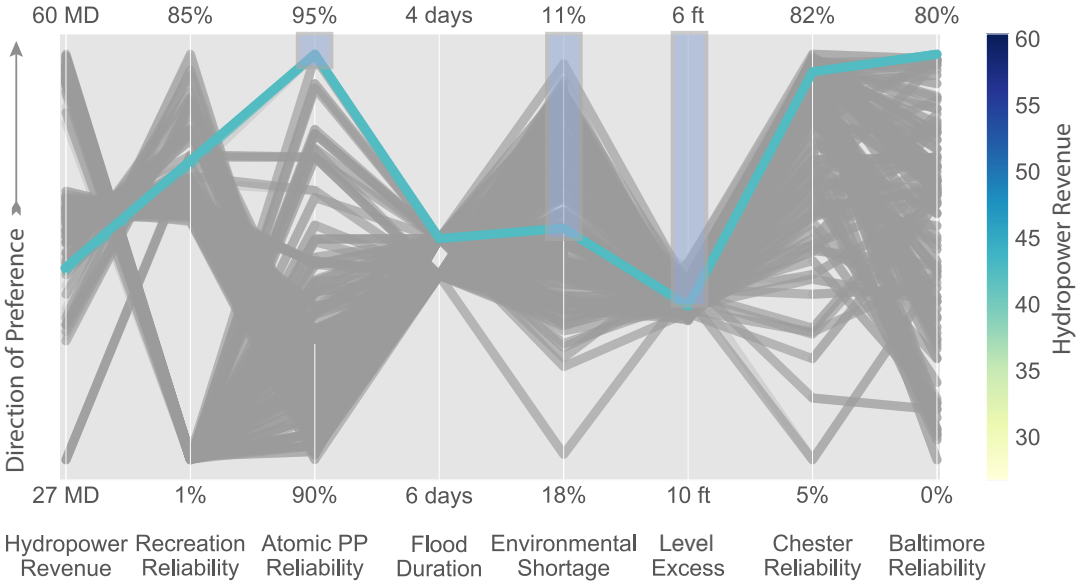

Figure 6.6: A compromise solution is found for the system through prioritizing performance across critical objectives. High performance solutions are required for critical criteria with nuclear power plant reliability above 94%, environmental shortage index below 14% and level excess below 8.5 ft. The compromise policy also achieves high performance for urban water supply for Chester and Baltimore, and attains reasonable hydropower revenue amounting for 43 MD/year. Recreation is somewhat compromised with 65% reliability.

The selection of a compromise solution also included acceptable performance for flooding level excess. Limited performance ranges achieved collectively for this objective restricted the selection of a suitable compromise. The policy with best level excess value that simultaneously met our previous performance restrictions was of 8.5 ft, equivalent to a total freeboard of 2.5 ft. This criterion is included in our selection since it captures dam safety. Performance for this objective is critical, representing a measure of the dam's structural damage during severe flood events; ultimately it would capture dam overtop and potential dam breach. Figure 6.6 illustrates the resultant compromise policy.

This policy emphasizes high performance for the aforementioned critical objectives but it also achieves high reliability for urban water supply, almost at the best-attained value for both Chester and Baltimore. It also achieves reservoir level targets for recreation with 65% reliability while hydropower revenue stays close to 43 MD per year. Our compromise policy emerges from a highly risk averse and constrained formulation. However, Section 6.4.3 more carefully uncovers the benefits of structuring a highly risk averse formulation and more broadly diagnose the performance of our worst first percentile compromise solution.

6.4.3 Expected Annual Performance

Figure 6.7 unfolds the performance of the compromise policy selected in Section 6.4.2. Note that in the previous results, the solutions shown in the parallel axis coordinates in Figures 6.5 and 6.6 are the result of a worst first percentile aggregation. Alternatively, Figure 6.7 disaggregates the individual annual results for the compromise Conowingo reservoir control policy. The figure graphically depicts the expected annual performance for 100,000 hydrologic traces as opposed to showing only the aggregated worst first percentile objective values. In Figure 6.7, the synthetic streamflow is ranked from wet to dry and for each trace the annual expected value for each objective is reported. Each colored line represents annual expected performance for one year of the 100,000 simulated years. Each column corresponds to a different objective, the top of each column states the lower and upper annual expected value ranges achieved by each objective across all 100,000 synthetic streamflow years. The color scale indicates the percent of the best objective value achieved across different hydrologic scenarios;

a completely peach-colored column would mean that 100% of the best objective value was achieved across all the simulated years. These results confirm high performance for our critical objectives. Consistent with the focus on the brushed performance requirements in Figure 6.6, performance close to ideal is achieved by the nuclear power plant, environmental shortage index and level excess across a broad suite of hydrologies in Figure 6.7.

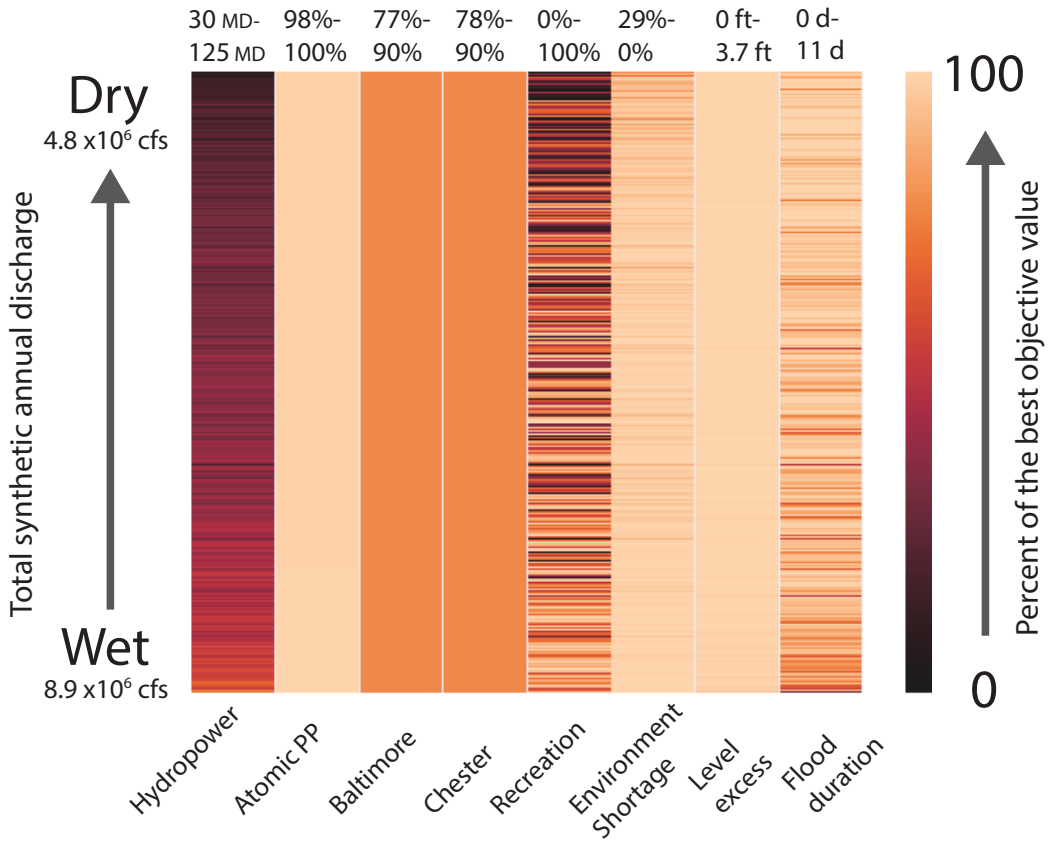

Figure 6.7: A risk neutral perspective showing expected annual performance for the compromise policy. The policy is re-evaluated across 100,000 hydrologic traces, which are ordered from wet to dry, and the corresponding average performance is calculated across objectives for each simulated year. The color scale is the percent of the best objective value. Where the lowest and highest value attained for each objective is shown at the top of each column.

Moreover, the expected value for urban water supply reliability is 90% for 80% of the scenarios tested, note that the lower ranges go as low as 77% and 78% for Baltimore and for Chester, this is far from the performance variability captured in our previous tradeoff analysis in Figure 6.5 with observed performance as low as 0% and 5%. The recreation objective; however, still shows variable performance when measuring expected values at an annual time-scale. As discussed earlier, the recreation objective is very sensitive to the small time frame in which critical storage is required. This objective also provides a sense of the state of the reservoir at the time when the highest multi-sectorial demands take place (i.e. downstream releases for fish passage and increased power demand). Performance in this objective may often be neglected and preference may be set on other pressing demands for the system. Variable performance is also observed for hydropower production within the annual expected value aggregation. However, the performance ranges lean towards a higher revenue, going up to 125 MD/year.

An alternative way of visualizing performance across our critical objectives is shown in the cumulative distribution plots in Figure 6.8. In this plots, high performance is shown towards the right for maximization objectives while high performance points towards the left for minimization objectives. The y-axis reflects the percent of the hydrologic scenarios in which a certain performance is achieved. Figure 6.8 panel (a) shows near ideal performance for the Nuclear Power Plant with reliability ranging from 98 to 100%. This results from the re-evaluation of the compromise solution across 100,000 synthetic hydrologies. In terms of the level excess objective shown in Panel (b), this objective stays very close to the desired reservoir level, achieving 0 deviation from the desired 109.2 ft during high streamflow events. The worst case for this objective still achieves

a freeboard of 7 ft.

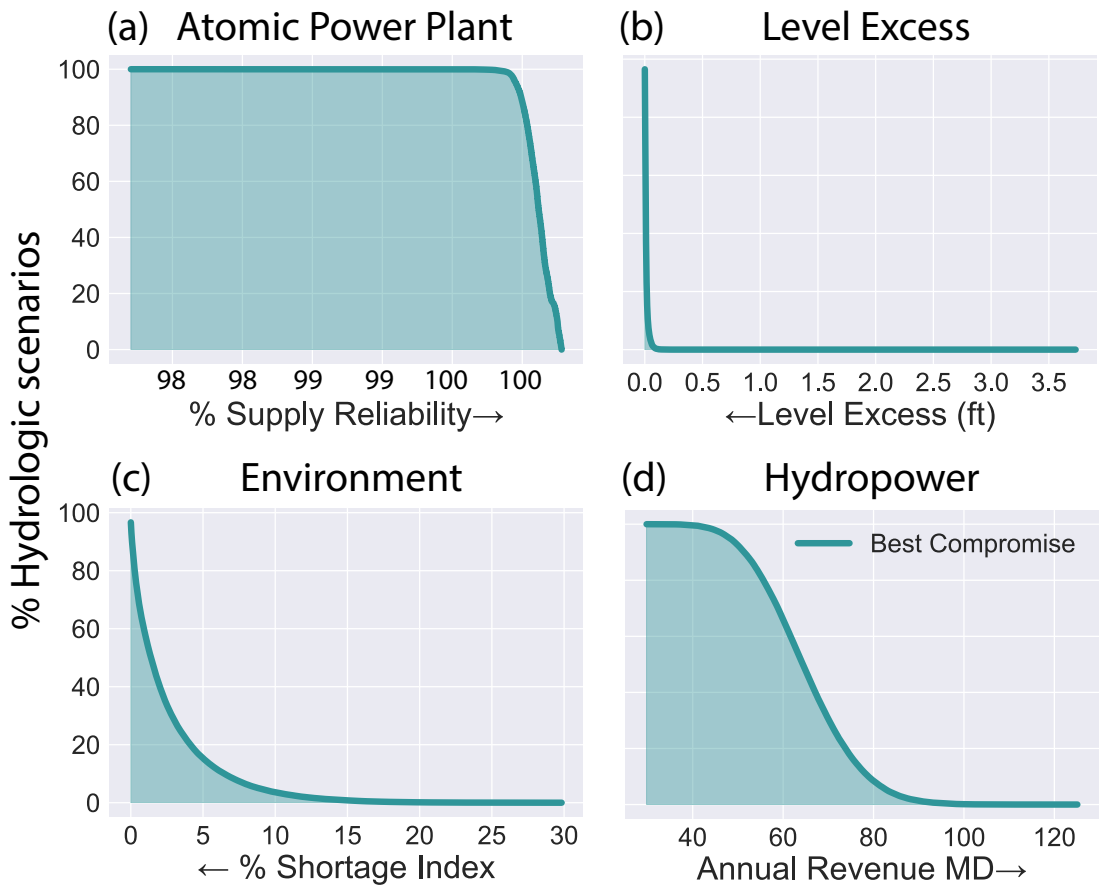

Figure 6.8: The cumulative distribution functions show the compromise control policy's performance across 100,000 synthetic traces. The desired values point towards the right for maximization objectives (panels a and b), whereas minimization objectives show preferred values towards the right (panels b and c). Panels a-c show near ideal performance for critical objectives while attaining good hydropower revenue in panel d across all the simulated years.

Panel (c) in Figure 6.8 shows minimum deviation from the required environmental flows and the actual releases for the majority of the hydrologic traces. Avoiding deviations for this objective will support a healthy ecosystem downstream and favor fishery resources. Finally, Panel (d) shows performance for hy-

dropower revenue, which ranges from 30 MD for very dry years up to 125 MD; our compromise solution achieves reasonable revenue for the majority of the streamflow cases used for its re-evaluation. This analysis shows high-expected annual performance for critical objectives while sustaining good hydropower revenue. The critical objectives seem to be less sensitive to the hydrology and perform well across the annual expected value for multiple simulated years. This outcome shows that defining a highly risk averse formulations can guarantee better average performance across multiple years.

6.4.4 Policy Dynamics

In this section, the resultant changes in control policies' dynamic behaviors are diagnosed in greater detail as a function of alternative objective performance preferences for the Conowingo reservoir system. These preferences are abstracted in Figure 6.9 by the best performing policies for hydropower revenue, nuclear power plant reliability and level excess. The compromise solution selected in section 6.6 is also included relative to the other performance baselines.

Each of these control policies was re-evaluated across 100,000 synthetic inflows to estimate the time varying probability density functions (PDFs) of the water level at Conowingo over time illustrated in Figure 6.10. The figure shows these estimates in log space for each of the solutions, with high probabilities shown in red, moderate probabilities yellow, and low probabilities blue. A solid line is drawn at the maximum height of 120 ft for the dam (i.e., the physical breach or overtopping level). The red high probability density region is maintained above 108 ft across all of the policies, regardless of preference. All of the

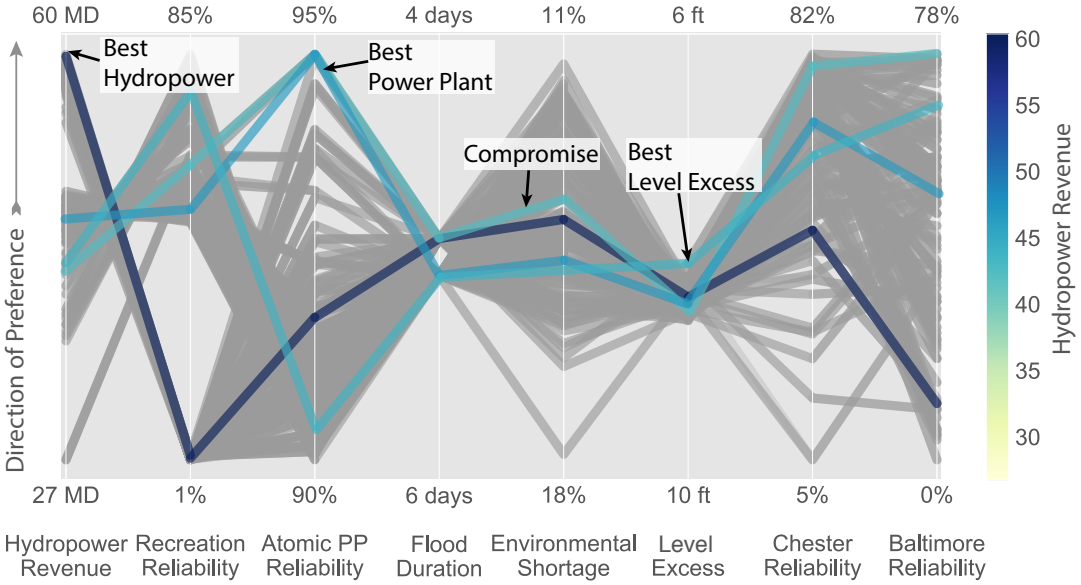

Figure 6.9: Capturing preference across sectors. Selecting the best policies for hydropower revenue, nuclear power plant, level excess and the compromise solution to analyze policy dynamics across preferences

policies stay close to the normal water level established by the SRBC target of 108.5 ft (Swartz, 2006). This level is suitable for sustaining water demands in the LSRB. As reference, Figure 6.1 shows the minimum levels required to meet the water demands as well as the level required for recreation. All the policies were able to keep water levels above these critical thresholds over time.

In all panels of Figure 6.10, low probability spikes for flood-based overtopping events (i.e., dark blue points) occur at the end of January during ice jams and from March through April when snowmelt and rain events are common. Once streamflow subsides at the end of the snowmelt season, the water levels begin to fall and lower levels are maintained throughout the dry season from June through September. This pattern is consistent for all the policies across multiple simulated years. The dynamics of our policies are capturing observed

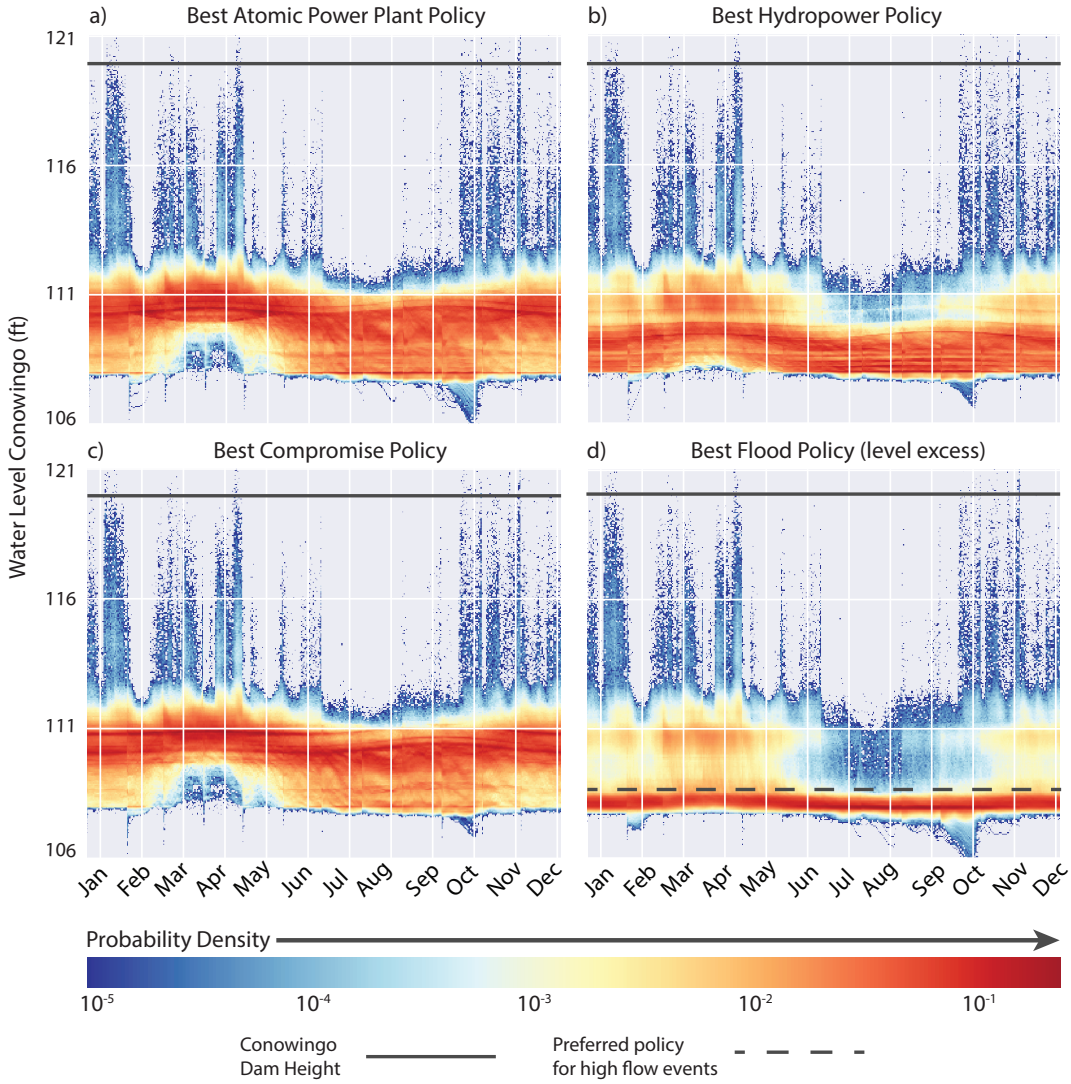

Figure 6.10: Probabilistic trajectories of the water level at Conowingo. The solid line shows the maximum height the the dam at 120 ft. High density is indicated in dark red while low density is shown in dark blue. All policies show similar shape driven by the streamflow trajectories for the system. The compromise solution in panel (c) is similar to the nuclear power plant policy in panel (a), indicating a dominant preference towards nuclear power plant when selecting a compromise. The hydropower policy in panel (b) closely follows the dynamics of peak energy demand, while the level excess policy maintains the level below the desired 109.2 ft shown by the dashed line.

hydrologic events in the LSRB region. Figure 6.10, shows that the operations for the best Nuclear Power Plant policy (panel a), the water level at Conowingo is maintained between 109 ft and 111 ft in a majority tested cases. These levels are well above the critical level established for Peach Bottom of 103.5ft. Our risk averse worst first percentile formulation is strongly shaped by reliably supplying cooling water demands from the nuclear power plant, generally keeping water levels at least 5.5 ft above the critical level. In Figure 6.10, operations that favor hydropower production (panel b), ensure a high probability of sustaining water levels between 108 ft and 110 ft, this range remains stable throughout the year. According to the SRBC, if the reservoir level reaches 104 ft, operations at Muddy Run Pumped hydro facility need to be shut down. This would represent a strongly negative loss of potential hydropower revenue and less flexibility in meeting the dynamic peak electricity demands on the area's regional power grid. At 1,070 megawatts Muddy Run is 50% larger than the installed capacity at the Conowingo reservoirs generating station. Furthermore, the hydropower policy reflects seasonal operations at Muddy Run, with moderate probability of high water elevation in the spring and fall when the demand for power is low. In contrast, panel b of Figure 6.10 shows that the maximum hydropower solution yields larger probabilities for higher levels during the winter and summer when the demand for power is at peak. The hydropower policy achieves a level of at least 4 ft above Muddy Runs critical level during the time when high power demands encounter low flows. In Figure 6.10, the storage trajectories for the compromise solution (panel c) closely resemble those of the nuclear power plant. This result is consistent with our selection of a compromise solution, which was dominated by performance constraints for the nuclear power plant. The compromise policy has high probability of remaining at a reservoir

level between 110 ft and 111ft; this level suggests high probability of meeting all the systems water demands. The objectives that are impacted by low flows agree with keeping high reservoir elevation; this could cause conflict with the flood policy (panel d) which attempts to keep a level below 109.2 ft, illustrated with the dotted line. The flood policy agrees with the reservoirs needs to keep level low and free space to anticipate large inflow, especially during snowmelt and hurricane season. The policy also shows a dip in September potentially driven by previous hurricanes during that time. Across all policies, the general shape of the time-varying PDFs is similar, the probabilities differ for certain preferences, such as hydropower, but the most marked difference is observed for the flood policy, which keeps the reservoir level stably below the desired level for high inflow events. It can be inferred that the selected preferences and our compromise solution achieve good performance during droughts, keeping reservoir levels far above the critical values stated by the SRBC. Conversely, during high streamflow all policies reflect a small probability of overtopping during challenging wet scenarios.

6.4.5 How Does the Discovered Compromise Policy Compare to Historical Operations?

After diagnosing how changes in preference manifest in the dynamics of candidate Conowingo reservoir control policies in the previous section, this section explores how the illustrative compromise solution's operations compare to historical operations using challenging historical and synthetically generated hydrologic extremes. For our *ex post* analysis, historical operations where ab-

stracted as described in the Methods in Section 6.4.5. Giuliani et al. (2014a) previously proposed an implicit optimization-based approach to capture historical operations of the Conowingo dam by assuming that the operators would meet urban water supply and would maximize hydropower revenue. Alternatively, in this study, gaged historical downstream releases where exploited to identify a control policy that closely approximates the actual historical release decisions. For this purpose, the Nash-Sutcliffe Efficiency Index (NSE) is maximized over the validation period of 1968-2016 using both upstream and downstream gages. The multi-master Borg MOEA achieved a value of 85% for the worst year approximation within 214,000 function evaluations.

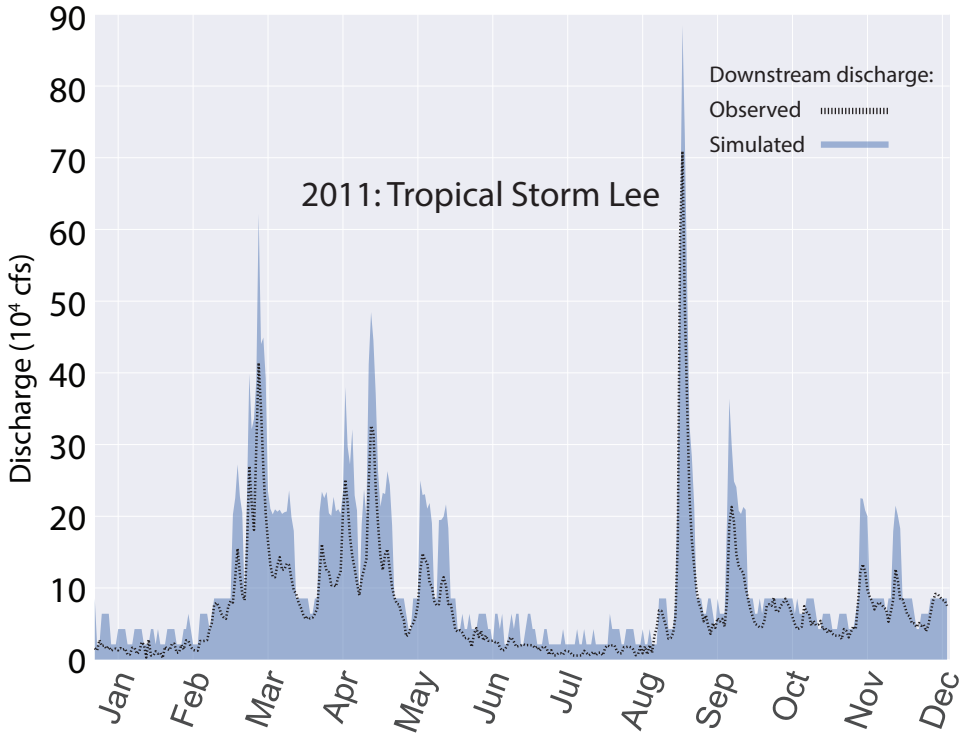

Figure 6.11: Performance for the Nash- Sutcliffe approximation of historical policy for the validation period of 1968-2016. A close-up of the year of interest 2011; the year when Tropical Storm Lee stroke. The NSE suitability is confirmed for this extreme event.

Figure 6.11 shows the results of this approximation for the validation period. Zooming into 2011 when Tropical Storm Lee stroke, which allows to assess the suitability of this approximation for the year of interest. The approximation was relatively accurate for this year and appropriate for our comparison of the downstream tradeoffs.

This style of approximation is limited by the availability of historical time series. Fortunately, there is a USGS gage immediately downstream of the reservoir with an extensive historical record available. A similar time-series is not available for the other releases; therefore, our tradeoff comparison for this analysis is only focused on the objectives that are impacted by the downstream releases such as hydropower production, environmental flows and releases through flood control gate, this is why only a subset of three downstream objectives is shown. The historical policy and compromise solution are re-evaluated under a dry and wet year and under a 100,000-member ensemble, independent to the one used for the re-evaluation of the policies in the previous sections. For the re-evaluation across single historical years, the worst-case value is used, that would mean that we take the minimum hydropower revenue and maximum flood duration and shortage index for that year. For the re-evaluation across an independent 100,000 set, again, the worst first percentile value is captured. These tradeoffs are shown in Figure 6.12, were the direction of preference is upwards.

Figure 6.12 panel (a) shows the downstream tradeoffs for the historical wet year were Tropical Storm Lee (2011) was captured, the compromise policy sacrifices performance in hydropower revenue, potentially as a result of implicitly capturing the requirement to keep reservoir levels low to accommodate high

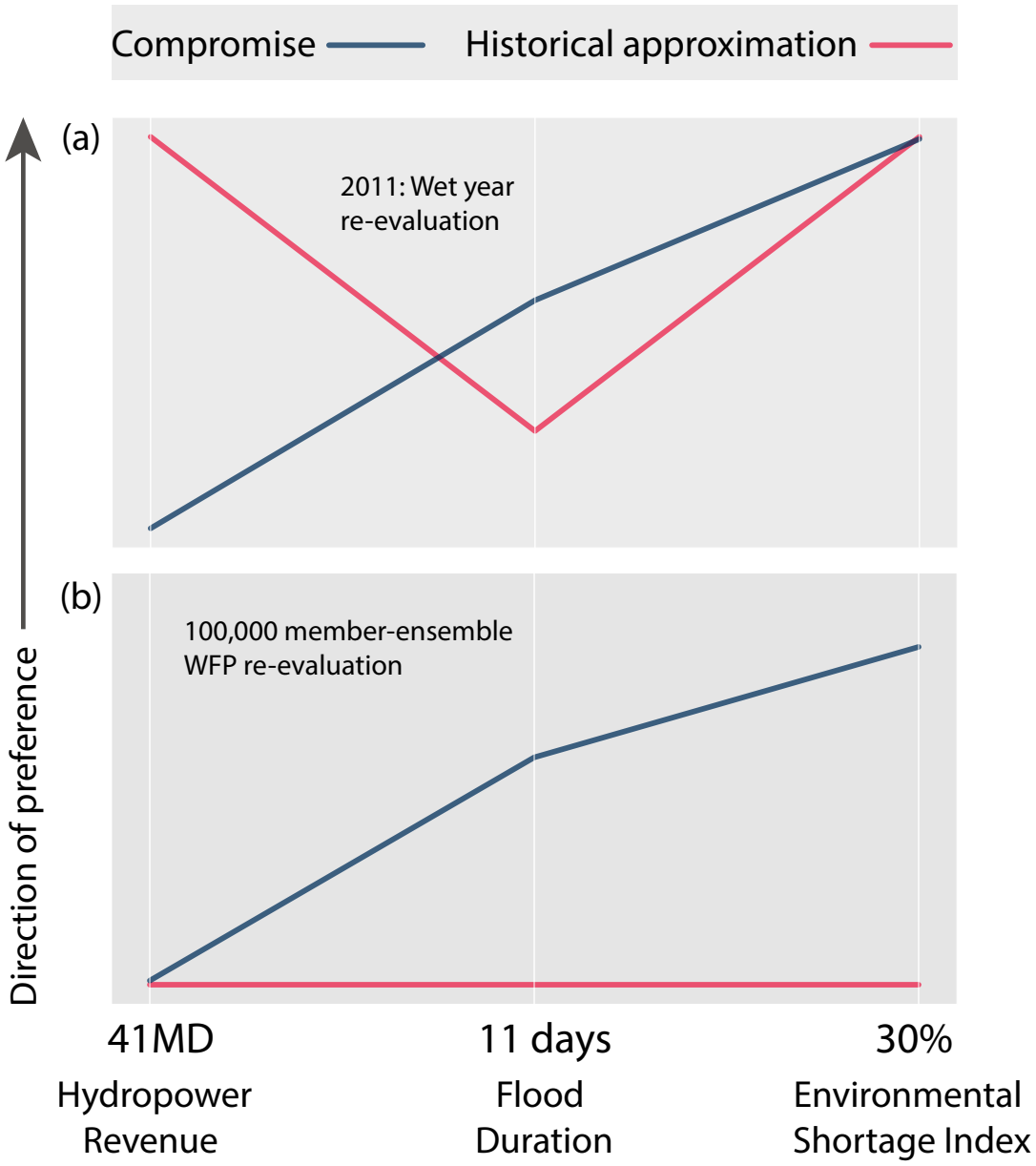

Figure 6.12: Tradeoffs of the compromise solution and the approximated historical policy attained by re-evaluating within a dry year, a wet year and through a larger sample with 100,000 independent traces. Panels a and b show that the *ad hoc* historical policy performs well for the specific wet event but it collapses in panel b when re-evaluated under a broader suite of hydrologies.

inflow. The compromise solution achieves prompt recovery from flashy downstream releases, within 4.6 days as opposed to 8 days for the historical policy, both policies achieve close to 0% environmental shortage index, meeting this goal was not problematic due to large water availability for that year. The policies were re-evaluated under a broader synthetic hydrology in panel (b). They both achieve similar hydropower revenue, but they start to diverge across the remaining objectives. The compromise policy outperformed the historical policy for flood duration and environmental shortage index. Even for this smaller subset of objectives, the historical policy is outperformed by the compromise policy when it is tested under more challenging hydrologies. This provides some indication of how the tradeoffs would behave when the hydrology does not follow historical trend when using an *ad hoc* policy that is solely suited for specific emergency conditions. The collapse of the historical policy when evaluated out of sample provides intuition about performance for the broader multisectoral demands under this policy, providing an indirect indication that other demands would potentially suffer as well. In contrast, the compromise policy generalizes well under a larger sample and its structure maintains implicit bias towards high water supply reliability.

6.5 Conclusions

This study advances the discovery of policies that balance multi-sectorial demands and minimize flood risk in the LSRB. Operations at Conowingo are key to meet the LSRB's demands for urban water supply for Baltimore, MD, Chester, PA, and cooling water for Peach Bottom nuclear power plant. The main purpose of the dam is hydropower production but operations are subject to guaranteeing downstream releases to protect fishery resources and upon maintaining levels for recreation. Tensions between these multi-sectorial demands emerge during severe drought conditions. Additionally, the dam represents the last flood control barrier between the Susquehanna River and several towns lying downstream from the Conowingo dam. The need to free storage for large inflow is in direct conflict with keeping storage levels high to meet water demands. To the best knowledge available, this is the first study that attempts to define policies to improve operations for the LSRB for both hydrologic extremes simultaneously. Risk-averse EMODPS policies are able to identify robust control policies when challenging extremes strike the system. In particular, minimizing the worst first percentile of the annual worst-case results in stable objective values that are able to reduce damages downstream and improve the capability of reliably meeting water demands. This abstraction of the Conowingo system accurately represents the LSRB's concerns. The resulting many objective stochastic control problem is high dimensional with operational and performance constraints. Discovery of feasible solutions that simultaneously met both performance constraints was possible due to the cooperative and adaptive search mechanisms of the multi-master Borg MOEA. It enabled an effective exploration of a difficult search space stemming from this challenging problem

structure. The resulting tradeoffs reflected a reduced feasible operating space with narrow performance ranges for critical objectives. In spite of methodological challenges when finding Pareto approximate policies for this verse of the Conowingo system, a compromise solution stroke high performance for critical objectives while attending the other demands. The compromise policy exposes key benefits of robust and constrained optimization 1) performance biases established in the optimization withstand re-evaluation under a broader hydrologic set and directly distinguish alternatives that would be relevant in practice 2) the tradeoffs for the system remain invariable when tested under more pronounced extremes and 3) high expected performance is certain when revealing annual performance across worst first percentile policies. By contrasting historical operations, it is observed that *ad hoc* release policies can result in rushed decisions and lead to performance losses across multiple demands. Moreover, the historical operations do not generalize well and completely collapse when evaluated under more challenging droughts and floods. In the contrary, the EMODPS approach helped define compromises that balance multi-sectorial demands while minimizing downstream and dam risks by embedding multiple challenging droughts and floods in the process. Additionally, the EMODPS policies generalized well and showed robust performance across more extreme hydrologic conditions. The control problem explored in this study is representative of the contextual and mathematical difficulties that are faced in a broad range of global multi-purpose systems challenged by the timings, severity and duration of hydrologic extremes. This study tested the ability to discover and appropriately manage water and energy supply, as well as the dam's capability to protect from floods. Nonetheless, this problem formulation represents an initial step in the attempt to improve management in the LSRB. Future work will be

focused on further improving operations by incorporating inflow information in addition to storage and time index into our policy design. Accounting for inflows would capture the use of forecast in our release decisions. Shaping our control policy with the previous days inflow will help anticipate operations and make better use of information. For an foreseen hurricane event, for example, a rational operator would empty the reservoir to have enough storage to contain high streamflow; conversely, releases would be minimized when expecting low inflows.

CHAPTER 7

CONTRIBUTIONS AND FUTURE WORK

7.1 Conclusions and Contributions

Research in the water resources field plays a pivotal role for defining the tools in which to aid in the discovery of management decisions that foster sustainability, robustness and reliability (Hashimoto et al., 1982b) in stressed systems. Advances in multi-objective optimization (Nicklow et al., 2010) have the potential to improve our ability to manage water resources across sectors that now experiencing a mix of acute and persistent stressors from climate and regional socio-economic changes. Uncertain extremes (e.g., floods and droughts) and growing human pressures when abstracted appropriately yield mathematical properties that challenge traditional optimization methods. In the attempt to solve more challenging water resources problems, the solution techniques also need to evolve and be clearly benchmarked in their strengths and weaknesses. This has motivated new developments in the multiobjective evolutionary optimization field, with multi-objective evolutionary algorithms MOEAs at the center of discovering tradeoffs for a number of water resources applications (Reed et al., 2013a; Kasprzyk, 2013). These studies are strongly promoting a shift towards *a posteriori* decision support for planning and management problems with the capability of removing the decision makers preconceptions by allowing the explicit representation of the systems tradeoffs for evolving or rival framings of problems (i.e., uncertainty and contention in how we should formulate the problems themselves) (Quinn et al., 2017a). Multi-purpose reservoir control represents a critical research area that remains reluctant to adopt inno-

vative techniques for new control policy design and re-operation of currently stressed reservoir systems. To this end, this dissertation contributes the most rigorous assessment MOEAs' capability to support robust multi-purpose reservoir policy design. The contributions of this work are in response to the needs directly identified by the fields most widely cited review by Nicklow et al. (2010) that specifically recognized the importance of rigorously characterizing emerging optimization tools to bridge existing research to practice while also broadening the scope of water systems' problems that can be addressed. The analysis focuses on the Lower Susquehanna River Basin (LSRB) system where multiple competing objectives for hydropower production, urban water supply, recreation and environmental flows need to be balanced. Chapter 6 broadened this representation of the system to also capture flooding concerns in the LSRB. The control problem explored in this study is representative of the contextual and mathematical challenges that are faced in a broad range of global multi-purpose systems challenged by multiple competing demands, uncertainty and performance constraints. Chapter 4 demonstrated that advances in MOEA design helped the discovery of suitable control policies for multi-purpose reservoir operations. The most valuable advances are those that allow a stable and bounded search for high-dimensional problems, have adaptive use of operators to search a varied decision space, and guarantee progress and improvements to the Pareto set throughout the search. MOEAs equipped with these traits are capable of reliably and effectively finding control policies that balance conflicting tradeoffs for multi-purpose reservoir control using Evolutionary Multi-objective Direct Policy Search (EMODPS). The potential value of MOEA solution techniques coupled with direct policy search lies in the capability of solving increasingly complex multi-purpose and potentially multi-reservoir control problems. Chapter

5, building off the results from Chapter 4, demonstrates the benefits of cooperative parallel MOEA architectures to reliably and effectively find many objective control policies when the system is subject to uncertainty and computational constraints. Hydrologic uncertainty was captured through different problem formulations, each with an increasing number of Monte Carlo samples embedded in the simulation. This analysis helped capture effects in the multi-objective search when the system is subject to approximate or noisy Monte Carlo (MC) function evauations. The analysis emphasizes the use of innovative cooperative search techniques implemented on parallel computing architectures to overcome computational constraints resulting from expensive function evaluations when embedding a large set of MC evaluations. Coordinated optimization benefitted the search dynamics helping overcome function evaluation constraints, tackle problem difficulty and reliably discovering high quality tradeoffs for the river basin. The insights from this chapter should enable water resources analysts to devote computational efforts towards representing reservoir systems more accurately by capturing uncertainty and multiple demands when properly using parallel coordinated search. Finally, this style of optimization will become more useful as the as computational capabilities continue to grow, increasing opportunities to utilize parallel metaheuristics in water resources applications. Chapter 6 expanded the multi- purpose reservoir control problem to also capture flood protection in the LSRB. A risk-averse EMODPS formulation contributed to the definition of policies to improve operations for the LSRB during droughts and floods. This problem structure yielded stable objective values that were able to reduce damages downstream and improve the capability of reliably meeting water demands. However, a more accurate representation of the LSRB's concerns, resulted in a many objective, stochastic, high dimensional con-

trol problem with operational and performance constraints. Discovery of feasible solutions that simultaneously met the performance constraints was possible due to the cooperative and adaptive search mechanisms of the multi-master Borg MOEA discussed in Chapter 5. Chapter 6 confirms the algorithm's ability to discover and appropriately manage water and energy supply, as well as the dam's capability to offer flood protection. This dissertation has contributed to rigorously assess the validity of recent advances in multi-objective decision support to design robust multi-purpose reservoir operations.

7.2 Future Work

Future work should focus on information selection and policy formulation to discover opportunities to build more resilient and robust policies for the LSRB as well as other river basins confronting similar multi-sectoral demands that must withstand increasingly more extreme floods and droughts. The three studies in this dissertation highlighted the potential of EMODPS in formulating multi-objective policies to build more robust systems; future efforts should be focused on targeting remaining challenges in policy approximation for this approach.

7.2.1 Information Selection and Policy Formulation

Central to the ability of policy design of the EMODPS is balancing the amount of information used in their design. The problem can be succinctly summarized as determining what inputs and how many of them to include during policy

design in order to best balance information gain against excessive specification (Galelli et al., 2014). This dissertation used reservoir level and time index to shape the reservoir control policies, inflow was not considered and could potentially improve the suitability of the discovered control policies. Even if there is more information available to the system, there is no clear way on how to decide the extent of information we need to shape the policies, information should help appropriate design of the system's operations but should not get in the way of defining adaptable policies that are able to perform well if the assumed conditions change. This challenge can also apply to other aspects of the policy structure such as number of radial basis functions and number of outputs. These questions are to an extent dependent upon the system's goals and their mathematical abstraction. Currently, there is no clear basis on how to choose the parameters of the global approximators other than empirical rules, also defining which inputs are most appropriate to include will also depend on the system objectives and how they're quantified. There is much opportunity for exploration in this area using metrics from information theory to attempt to balance benefits of information versus overfitting, and to achieve good balance for the specification of parameters of the radial basis policies.

7.2.2 Capturing Non-stationarity and Impacts to Increasing Water Demands in the Lower Susquehanna River Basin system

This dissertation focused on evaluating the impacts of the LSRB system under stationary hydrologic uncertainty. However, the capability of meeting all the

basin needs may be potentially impacted by changes in the hydrologic regimes. Changes in the mean annual streamflow and shifting snowmelt for instance, can severely impact the system's reliability due to shifts on the timings of critical demands and water availability. Testing the system under non-stationary hydrology can help understand the system's vulnerability to these changes. Future research efforts should also focus on extensions that consider broader envelopes of uncertainty to encompass societal challenges, climate change and capacity restrictions in the Conowingo reservoir. The LSRB is experiencing well-known expansion of its water supply capabilities, adding new urban water supply and energy requirements in the area. Climate change could generate changes in the streamflow patterns as discussed earlier, and finally the storage capacity is been severely impacted by sediment trapped behind the dam. To gage the impacts of these uncertain factors, the LSRB system would benefit from sensitivity analysis (Lempert et al., 2008) to determine which conditions control the robustness of the system and inform the combination of factors that can contribute to the system's failure to meet its multi-sectorial demands. This analysis could be used in future work to inform when policies should be re-designed (Groves et al., 2014) or on mitigating the negative effects of the controlling factors (Herman et al., 2014).

BIBLIOGRAPHY

Alba, E., Luque, G., and Nesmachnow, S. (2013). "Parallel metaheuristics: recent advances and new trends." *International Transactions in Operational Research*, 20(1), 1–48.

Basdekas, L. (2014). "Is multiobjective optimization ready for water resources practitioners? utility's drought policy investigation." *Journal of Water Resources Planning and Management*, 140(3), 275–276.

Bellman, R. (1957). "A markovian decision process." *Report no.*, DTIC Document.

Beyer, H.-G. and Sendhoff, B. (2007). "Robust optimization–a comprehensive survey." *Computer methods in applied mechanics and engineering*, 196(33), 3190–3218.

Brekke, L. D., Maurer, E. P., Anderson, J. D., et al. (2009). "Assessing reservoir operations risk under climate change." *Water Resour. Res.*, 45(4), n/a–n/a.

Brown, C. M., Lund, J. R., Cai, X., et al. (2015). "The future of water resources systems analysis: Toward a scientific framework for sustainable water management." *Water resources research*, 51(8), 6110–6124.

Burke, E. K., Gendreau, M., Hyde, M., et al. (2013). "Hyper-heuristics: A survey of the state of the art." *Journal of the Operational Research Society*, 64(12), 1695–1724.

Cantu-Paz, E. (2000). *Efficient and accurate parallel genetic algorithms*, Vol. 1. Springer Science & Business Media.

Castelletti, A., Antenucci, J., Limosani, D., Quach Thi, X., and Soncini-Sessa, R. (2011). "Interactive response surface approaches using computationally intensive models for multiobjective planning of lake water quality remediation." *Water Resources Research*, 47(9).

Castelletti, A., Pianosi, F., Quach, X., and Soncini-Sessa, R. (2012a). "Assessing water reservoirs management and development in northern vietnam." *Hydrol. Earth Syst. Sci*, 16, 189–199.

Castelletti, A., Pianosi, F., and Restelli, M. (2013). "A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run." *Water Resources Research*, 49(6), 3476–3486.

Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2008). "Water reservoir control under economic, social and environmental constraints." *Automatica*, 44(6), 1595–1607.

Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2012b). "Stochastic and robust control of water resource systems: Concepts, methods and applications." *System Identification, Environmental Modelling, and Control System Design*, 383–401.

Castelletti, A., Pianosi, F., Soncini-Sessa, R., and Antenucci, J. P. (2010). "A multiobjective response surface approach for improved water quality planning in lakes and reservoirs." *Water Resour. Res.*, 46(6), n/a–n/a.

Castelletti, A., Yajima, H., Giuliani, M., Soncini-Sessa, R., and Weber, E. (2014a). "Planning the optimal operation of a multioutlet water reservoir with water quality and quantity targets." *Journal of Water Resources Planning and Management*, 140(4), 496–510.

Castelletti, A., Yajima, H., Giuliani, M., Soncini-Sessa, R., and Weber, E. (2014b). "Planning the optimal operation of a multioutlet water reservoir with water quality and quantity targets." *Journal of Water Resources Planning and Management*, 140(4), 496–510.

Chankong, V. and Haimes, Y. (1983). *Multiobjective Decision Making: Theory and Methodology*. North-Holland, New York.

Characklis, G., Kirsch, B. R., Ramsey, J., Dillard, K., and Kelley, C. T. (2006). "Developing portfolios of water supply transfers." *Water Resources Research*, 42.

Coello, C. A. C., Lamont, G. B., Van Veldhuizen, D. A., et al. (2007). *Evolutionary algorithms for solving multi-objective problems*, Vol. 5. Springer.

Coello, C. A. C., Veldhuizen, D. A. V., and Lamont, G. B. (2002). "MOEA parallelization." *Genetic Algorithms and Evolutionary Computation*, Springer Science $\mathplus$ Business Media, 293–320.

Coello Coello, C. A. (2007). *Evolutionary Algorithms for Solving Multi-Objective Problems*. Genetic and Evolutionary Computation. Springer, New York, 2 edition.

Cohon, J. and Marks, D. (1973). "Multiobjective screening models and water resource investment." *Water Resources Research*, 9(4), 826–836.

Cohon, J. and Marks, D. (1975a). "A review and evaluation of multiobjective programming techniques." *Water Resources Research*, 11(2), 208–220.

Cohon, J. L., Church, R. L., and Sheer, D. P. (1979). "Generating multiobjective trade-offs: An algorithm for bicriterion problems." *Water Resources Research*, 15(5), 1001–1010.

Cohon, J. L. and Marks, D. H. (1975b). "A review and evaluation of multiobjective programing techniques." *Water Resources Research*, 11(2), 208–220.

Crowl, L. A. (1994). "How to measure, present, and compare parallel performance." *IEEE Parallel & Distributed Technology: Systems & Technology*, 2(1), 9–25.

Cui, L. and Kuczera, G. (2005). "Optimizing water supply headworks operating rules under stochastic inputs: Assessment of genetic algorithm performance." *Water resources research*, 41(5).

Dariane, A. B. and Momtahen, S. (2009). "Optimization of multireservoir systems operation using modified direct search genetic algorithm." *Journal of Water Resources Planning and Management*, 135(3), 141–148.

Deb, K. and Agrawal, R. B. (1994). "Simulated binary crossover for continuous search space." *Report No. IITK/ME/SMD-94027*, Indian Institute of Technology, Kanpur.

Deb, K. and Goyal, M. (1996). "A combined genetic adaptive search (geneas) for engineering design." *Computer Science and Informatics*, 26, 30–45.

Deb, K. and Gupta, H. (2006a). "Introducing robustness in multi-objective optimization." *Evolutionary Computation*, 14(4), 463–494.

Deb, K. and Gupta, H. (2006b). "Introducing robustness in multi-objective optimization." *Evolutionary computation*, 14(4), 463–494.

Deb, K., Mohan, M., and Mishra, S. (2003). "A fast multi-objective evolutionary algorithm for finding well-spread pareto-optimal solutions." *Report No. 2003002*, Kanpur Genetic Algorithms Laboratory.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197.

Deb, K. and Sinha, A. (2009). "Solving bilevel multi-objective optimization problems using evolutionary algorithms." *Evolutionary Multi-Criterion Optimization*, Springer Science $\mathplus$ Business Media, 110–124.

Dougherty, D. E. and Marryott, R. A. (1991). "Optimal groundwater management: 1. simulated annealing." *Water Resour. Res.*, 27(10), 2493–2508.

Efstratiadis, A., Koutsoyiannis, D., and Xenos, D. (2004). "Minimizing water cost in water resource management of athens." *Urban Water Journal*, 1(1), 3–15.

Faber, B. A. and Stedinger, J. (2001). "Reservoir optimization using sampling sdp with ensemble streamflow prediction (esp) forecasts." *Journal of Hydrology*, 249(1), 113–133.

Fleming, P. J., Purshouse, R. C., and Lygoe, R. J. (2005). "Many-objective optimization: An engineering design perspective." *Evolutionary multi-criterion optimization*, Springer, 14–32.

Giuliani, M. and Castelletti, A. (2016). "Is robustness really robust? how different definitions of robustness impact decision-making under climate change." *Climatic Change*, 1–16.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., and P.M., R. (2015a). "Curses, tradeoffs, and scalable management: Advancing evolutionary multi-objective direct policy search to improve water reservoir operations." *ASCE Journal of Water Resources Planning & Management*.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., and Reed, P. M. (2015b). "Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations." *Journal of Water Resources Planning and Management*, 142(2), 04015050.

Giuliani, M., Herman, J., Castelletti, A., and Reed, P. (2014a). "Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management." *Water Resources Research*, 50(4), 3355–3377.

Giuliani, M., Herman, J. D., Castelletti, A., and Reed, P. M. (2013). "Many-objective reservoir policy identification and refinement to reduce institutional myopia in water management." *AGU Fall Meeting Abstracts*.

Giuliani, M., Herman, J. D., Castelletti, A., and Reed, P. M. (2014b). "Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management." *Water Resources Research*, 50(4), 3355–3377.

Giuliani, M., Pianosi, F., and Castelletti, A. (2015c). "Making the most of data: an information selection and assessment framework to improve water systems operations." *Water Resources Research*, 51(11), 9073–9093.

Goldberg, D. (2002a). "The design of innovation: Lessons from and for competent genetic algorithms kluwer academic publishers." *Norwell, Mass*.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

Goldberg, D. E. (2002b). *The Design of Innovation: Lessons from and for Competant Genetic Algorithms*. Kluwer Academic Publishers, Boston.

Gopalakrishnan, G., Minsker, B. S., and Goldberg, D. E. (2003). "Optimal sampling in a noisy genetic algorithm for risk-based remediation design." *Journal of Hydroinformatics*, 5(1), 11–25.

Groves, D., Bloom, E., Lempert, R., et al. (2014). "Developing key indicators for adaptive water planning." *Journal of Water Resources Planning and Management*, 0(0), 05014008.

Guariso, G., Rinaldi, S., and Soncini-Sessa, R. (1986). "The management of lake como: A multiobjective analysis." *Water Resources Research*, 22(2), 109–120.

Guo, X., Hu, T., Wu, C., Zhang, T., and Lv, Y. (2013). "Multi-objective optimization of the proposed multi-reservoir operating policy using improved nspso." *Water resources management*, 27(7), 2137–2153.

Hadka, D., Madduri, K., and Reed, P. (2013). "Scalability analysis of the asynchronous, master-slave borg multiobjective evolutionary algorithm." *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2013 IEEE 27th International*, IEEE, 425–434.

Hadka, D. and Reed, P. (2012a). "Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization." *Evolutionary Computation*, 20(3), 423–452.

Hadka, D. and Reed, P. (2012b). "Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization." *Evolutionary Computation*, 20(3), 423–452.

Hadka, D. and Reed, P. (2013a). "Borg: An auto-adaptive many-objective evolutionary computing framework." *Evolutionary Computation*, 21(2), 231–259.

Hadka, D. and Reed, P. (2013b). "Borg: An auto-adaptive many-objective evolutionary computing framework." *Evolutionary computation*, 21(2), 231–259.

Hadka, D. and Reed, P. (2014). "Large-scale parallelization of the borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems." *Environmental Modelling & Software*.

Hadka, D. and Reed, P. (2015). "Large-scale parallelization of the borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems." *Environmental Modelling & Software*, 69, 353–369.

Hadka, D., Reed, P. M., and Simpson, T. W. (2012). "Diagnostic assessment of the borg moea for many-objective product family design problems." *Evolutionary Computation (CEC), 2012 IEEE Congress on*, IEEE, 1–10.

Haimes, Y. Y. and Hall, W. A. (1977). "Sensitivity, responsivity, stability and irreversibility as multiple objectives in civil systems." *Advances in Water Resources*, 1(2), 71–81.

Hamarat, C., Kwakkel, J. H., Pruyt, E., and Loonen, E. T. (2014). "An exploratory approach for adaptive policymaking by using multi-objective robust optimization." *Simulation Modelling Practice and Theory*, 46, 25–39.

Hashimoto, T., Loucks, D. P., and Stedinger, J. R. (1982a). "Reliability, resiliency, robustness, and vulnerability criteria for water resource systems." *Water Resources Research*, 18(1).

Hashimoto, T., Stedinger, J. R., and Loucks, D. P. (1982b). "Reliability, resiliency and vulnerability criteria for water resource system performance evaluation." *Water Resources Research*, 18(1), 14–20.

Herman, J. D., Zeff, H. B., Lamontagne, J. R., Reed, P. M., and Characklis, G. W. (2016). "Synthetic Drought Scenario Generation to Support Bottom-Up Water Supply Vulnerability Assessments." *Journal of Water Resources Planning and Management*, 142(11), 04016050.

Herman, J. D., Zeff, H. B., Reed, P. M., and Characklis, G. W. (2014). "Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty." *Water Resources Research*, 50(10), 7692–7713.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Inselberg, A. (2009). "Parallel coordinates: Visual multidimensional geometry and its applicationss." *Spring science, New York*.

Iorio, A. W. and Li, X. (2008). "Improving the performance and scalability of differential evolution." *Simulated Evolution and Learning*, Springer, 131–140.

Kasprzyk, J. R. (2013). "Many objective water resources planning and management given deep uncertainties, population pressures, and environmental change." Ph.D. thesis, The Pennsylvania State University, The Pennsylvania State University.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., and Lempert, R. J. (2013a). "Many objective robust decision making for complex environmental systems undergoing change." *Environmental Modelling and Software*, 42, 55–71.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., and Lempert, R. J. (2013b). "Many objective robust decision making for complex environmental systems undergoing change." *Environmental Modelling & Software*, 42, 55–71.

Kasprzyk, J. R., Reed, P. M., Characklis, G. W., and Kirsch, B. R. (2012). "Many-objective de novo water supply portfolio planning under deep uncertainty." *Environmental Modelling and Software*, 34, 87–104.

Kern, J. D., Characklis, G. W., and Foster, B. T. (2015). "Natural gas price uncertainty and the cost-effectiveness of hedging against low hydropower revenues caused by drought." *Water Resources Research*, 51(4), 2412–2427.

Kirsch, B. R., Characklis, G. W., and Zeff, H. B. (2012). "Evaluating the impact of alternative hydro-climate scenarios on transfer agreements: Practical improvement for generating synthetic streamflows." *Journal of Water Resources Planning and Management*, 139(4), 396–406.

Kirsch, B. R., Characklis, G. W., and Zeff, H. B. (2013). "Evaluating the impact of alternative hydro-climate scenarios on transfer agreements: A practical improvement for generating synthetic streamflows." *Journal of Water Resources Planning and Management*, 139(4), 396–406.

Kita, H., Ono, I., and Kobayashi, S. (1999). "Multi-parental extension of the unimodal normal distribution crossover for real-coded genetic algorithms." *Congress on Evolutionary Computation*, 1581–1588.

Knowles, J. and Corne, D. (2002). "On metrics for comparing nondominated sets." *Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on*, Vol. 1, IEEE, 711–716.

Kollat, J. B. (2005). "The epsilon non-dominated sorted genetic algorithm ii: A highly effective multi-objective evolutionary algorithm for water resources applications." M.S. thesis, The Pennsylvania State University, The Pennsylvania State University.

Kollat, J. B. and Reed, P. M. (2005). "The value of online adaptive search: A performance comparison of NSGAII, epsilon-NSGAII and epsilon-MOEA." *EMO 2005: The Third International Conference On Evolutionary Multi-Criterion Optimization*, C. Coello Coello, A. Aguirre, and E. Zitzler, eds., number 3410 in Lecture Notes in Computer Science, Springer Verlag, 386–398.

Kollat, J. B. and Reed, P. M. (2006). "Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design." *Advances in Water Resources*, 29(6), 792–807.

Kollat, J. B. and Reed, P. M. (2007). "A computational scaling analysis of multi-objective evolutionary algorithms in long-term groundwater monitoring applications." *Advances in Water Resources*, 30(3), 335–353.

Kollat, J. B., Reed, P. M., and Maxwell, R. M. (2011). "Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics." *Water Resources Research*, 47.

Koutsoyiannis, D. and Economou, A. (2003). "Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems." *Water Resources Research*, 39(6).

Kukkonen, S. and Lampinen, J. (2005). "Gde3: The third evolution step of generalized differential evolution." *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, Vol. 1, IEEE, 443–450.

Labadie, J. W. (2004). "Optimal operation of multireservoir systems: State-of-the-art review." *Journal of Water Resources Planning and Management*, 130(2), 93–111.

Lall, U. and Sharma, A. (1996). "A nearest neighbor bootstrap for resampling hydrologic time series." *Water Resources Research*, 32(3), 679–693.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002a). "Combining convergence and diversity in evolutionary multiobjective optimization." *Evolutionary computation*, 10(3), 263–282.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002b). "Combining convergence and diversity in evolutionary multiobjective optimization." *Evolutionary computation*, 10(3), 263–282.

Lempert, R. J., Bryant, B. P., and Bankes, S. C. (2008). "Comparing algorithms for scenario discovery." *Report No. WR-557-NSF*, RAND.

Loucks, D. P., Van Beek, E., Stedinger, J. R., Dijkman, J. P., and Villars, M. T. (2005). *Water resources systems planning and management: an introduction to methods, models and applications*. Paris: Unesco.

Maass, A., Hufschmidt, M. M., Dorfman, R., et al. (1962). *Design of Water-Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning*. Harvard University Press, Cambridge.

Maier, H. R., Kapelan, Z., Kasprzyk, J., et al. (2014a). "Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions." *Environmental Modelling & Software*, 62, 271–299.

Maier, H. R., Kapelan, Z., Kasprzyk, J., et al. (2014b). "Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions." *Environmental Modelling & Software*, 62, 271–299.

Maier, H. R., Kapelan, Z., Kasprzyk, J., et al. (2014c). "Evolutionary algorithms

and other metaheuristics in water resources: Current status, research challenges and future directions." *Environmental Modelling & Software*, 62, 271–299.

Messac, A. and Mattson, C. A. (2002). "Generating well-distributed sets of Pareto points for engineering design using physical programming." *Optimization and Engineering*, 3(4), 431–450.

Mortazavi-Naeini, M., Kuczera, G., Kiem, A. S., et al. (2015). "Robust optimization to secure urban bulk water supply against extreme drought and uncertain climate change." *Environmental Modelling & Software*, 69, 437–451.

Moss, R., Fisher-Vanden, K., Delgado, A., et al. (2016). "Understanding dynamics and resilience in complex interdependent systems." *Washington, D.C.,U.S., Global Change Research Program Interagency Group on Integrative Modeling: 38*.

Müller, R. and Schütze, N. (2016). "Multi-objective optimization of multipurpose multi-reservoir systems under high reliability constraints." *Environmental Earth Sciences*, 75(18), 1278.

Nash, J. E. and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models part i—a discussion of principles." *Journal of hydrology*, 10(3), 282–290.

National Research Council (2009). *Informing Decisions in a Changing Climate*. National Academy Press.

Nicklow, J., Reed, P., Savic, D., et al. (2009). "State of the art for genetic algorithms and beyond in water resources planning and management." *Journal of Water Resources Planning and Management*, 136(4), 412–432.

Nicklow, J., Reed, P., Savic, D., et al. (2010). "State of the art for genetic algorithms and beyond in water resources planning and management." *Journal of Water Resources Planning and Management*, 136(4), 412–432.

Nowak, K., Prairie, J., Rajagopalan, B., and Lall, U. (2010a). "A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow." *Water Resources Research*, 46.

Nowak, K., Prairie, J., Rajagopalan, B., and Lall, U. (2010b). "A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow." *Water Resources Research*, 46(8).

Oliveira, R. and Loucks, D. P. (1997a). "Operating rules for multireservoir systems." *Water Resources Research*, 33(4), 839–852.

Oliveira, R. and Loucks, D. P. (1997b). "Operating rules for multireservoir systems." *Water resources research*, 33(4), 839–852.

Pachauri, R. K., Allen, M. R., Barros, V. R., et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC.

Pareto, V. (1896). *Cours D'Economie Politique*. Rouge, Lausanne.

Quinn, J. D., Reed, P. M., Giuliani, M., and Castelletti, A. (2017a). "Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems." *Water Resources Research*, 53(8), 7208–7233.

Quinn, J. D., Reed, P. M., Giuliani, M., and Castelletti, A. (2017b). "Rival framings: A framework for discovering how problem formulation uncertainties shape risk management tradeoffs in water resources systems." *(In Review) Water Resources Research*.

Reed, P. M. and Hadka, D. (2014). "Evolving many-objective water management to exploit exascale computing." *Water Resources Research*, 50(10), 8367–8373.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., and Kollat, J. B. (2013a). "Evolutionary multiobjective optimization in water resources: The past, present and future." *Advances in Water Resources*, 51, 438–456.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., and Kollat, J. B. (2013b). "Evolutionary multiobjective optimization in water resources: The past, present and future." *Advances in Water Resources*, 51, 438–456.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., and Kollat, J. B. (2013c). "Evolutionary multiobjective optimization in water resources: The past, present, and future." *Advances in water resources*, 51, 438–456.

Reed, P. M. and Kollat, J. B. (2013). "Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example." *Advances in Water Resources*, 56, 1–13.

Reyes-Sierra, M. and Coello, C. C. (2006). "Multi-objective particle swarm optimizers: A survey of the state-of-the-art." *International journal of computational intelligence research*, 2(3), 287–308.

Reynoso-Meza, G., Sanchis, J., Blasco, X., and García-Nieto, S. (2014). "Physical programming for preference driven evolutionary multi-objective optimization." *Applied Soft Computing*, 24, 341–362.

Rigg, J., Bebbington, A., Gough, K. V., et al. (2009). "The world development report 2009 'reshapes economic geography': geographical reflections." *Transactions of the Institute of British Geographers*, 34(2), 128–136.

Rosenstein, M. T. and Barto, A. G. (2001). "Robot weightlifting by direct policy search." *International Joint Conference on Artificial Intelligence*, Vol. 17, Citeseer, 839–846.

Roy, B. (1971). "Problems and methods with multiple objective functions." *Mathematical Programming*, 1(1), 239–266.

Roy, B. (1990). "Decision-aid and decision-making." *European Journal of Operational Research*, 45(2–3), 324–331.

Sheer, D. P. and Dehoff, A. (2009a). "Science-based collaboration: Finding better ways to operate the conowingo pond." *American Water Works Association. Journal*, 101(6), 20.

Sheer, D. P. and Dehoff, A. (2009b). "Science-based collaboration: Finding better ways to operate the Conowingo Pond." *Journal of the American Water Works Association*, 101(6), 20–22.

Simonovic, S. P. (1992). "Reservoir systems analysis: Closing gap between theory and practice." *Journal of Water Resources Planning and Management*, 118(3), 262–280.

Singh, A. and Minsker, B. S. (2008). "Uncertainty-based multiobjective optimization of groundwater remediation design." *Water resources research*, 44(2).

Soncini-Sessa, R., Weber, E., and Castelletti, A. (2007). *Integrated and participatory water resources management-theory*, Vol. 1. Elsevier.

Storn, R. and Price, K. (1997a). "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces." *Journal of Global Optimization*, 11(4), 341–359.

Storn, R. and Price, K. (1997b). "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces." *Journal of global optimization*, 11(4), 341–359.

Susquehanna River Basin Commission (2006). "Conowingo pond management plan." *Report No. 242*, Susquehanna River Basin Commission.

Susquehanna River Basin Commission (2008). "Consumptive use mitigation plan." *Report No. 253*, Susquehanna River Basin Commission.

Swartz, P. (2006). "Conowingo pond management plan." *Technical Report 242*, Susquehanna River Basin Commission.

Tanaka, S. K., Zhu, T., Lund, J. R., et al. (2006). "Climate warming and water management adaptation for california." *Climatic Change*, 76(3-4), 361–387.

Tang, Y., Reed, P. M., and Kollat, J. B. (2007). "Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications." *Advances in Water Resources*, 30(3), 335–353.

Tejada-Guibert, J. A., Johnson, S. A., and Stedinger, J. R. (1995). "The value of hydrologic information in stochastic dynamic programming models of a multireservoir system." *Water resources research*, 31(10), 2571–2579.

Teytaud, O. (2006). "How entropy-theorems can show that approximating high-dim pareto-fronts is too hard." *Bridging the Gap between Theory and Practice-Workshop PPSN-BTP*.

Trenberth, K. E. (2011). "Changes in precipitation with climate change." *Climate Research*, 47(1-2), 123–138.

Tsitsiklis, J. N. and Van Roy, B. (1996). "Feature-based methods for large scale dynamic programming." *Machine Learning*, 22(1-3), 59–94.

Tsoukalas, I., Kossieris, P., Efstratiadis, A., and Makropoulos, C. (2016). "Surrogate-enhanced evolutionary annealing simplex algorithm for effective

and efficient optimization of water resources problems on a budget." *Environmental Modelling & Software*, 77, 122–142.

Tsoukiàs, A. (2008). "From decision theory to decision aiding methodology." *European Journal of Operational Research*, 187(1), 138–161.

Tsutsui, S., Yamamura, M., and Higuchi, T. (1999). "Multi-parent recombination with simplex crossover in real coded genetic algorithms." *Genetic and Evolutionary Computation Conference (GECCO 1999)*.

Vamplew, P., Dazeley, R., Berry, A., Issabekov, R., and Dekker, E. (2011). "Empirical evaluation methods for multiobjective reinforcement learning algorithms." *Machine Learning*, 84(1-2), 51–80.

Van Veldhuizen, D. A. and Lamont, G. B. (1998a). "Evolutionary computation and convergence to a pareto front." *Late breaking papers at the genetic programming 1998 conference*, Citeseer, 221–228.

Van Veldhuizen, D. A. and Lamont, G. B. (1998b). "Multiobjective evolutionary algorithm research: A history and analysis." *Report no.*, Citeseer.

Vrugt, J. A. and Robinson, B. A. (2007). "Improved evolutionary optimization from genetically adaptive multimethod search." *Proceedings of the National Academy of Sciences*, 104(3), 708–711.

Ward, V. (2015). "Confronting tipping points: How well can multi-objective evolutionary algorithms support the management of environmental thresholds.

Ward, V., Singh, R., Reed, P., and Keller, K. (2015). "(in-revision). confronting tipping points: How well can multi-objective evolutionary algorithms support the management of environmental thresholds?." *Environmental Modelling & Software*.

Woodruff, M. J., Reed, P. M., and Simpson, T. (2013). "Many objective visual analytics: Rethinking the design of complex engineered systems." *Structural and Multidisciplinary Optimization*, 48(1), 201–219.

World Bank (2016). *High and Dry: Climate Change, Water and the Economy*. World Bank, Washington, DC.

Yeh, W. W.-G. (1985). "Reservoir management and operations models: A state-of-the-art review." *Water Resources Research*, 21(12), 1797–1818.

Zatarain-Salazar, J., Reed, P. M., Herman, J. D., Giuliani, M., and Castelletti, A. (2016). "A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control." *Advances in water resources*, 92, 172–185.

Zatarain Salazar, J., Reed, P. M., Herman, J. D., Giuliani, M., and Castelletti, A. (2016). "A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control." *Advances in Water Resources*, 92, 172–185.

Zatarain Salazar, J., Reed, P. M., Quinn, J. D., Giuliani, M., and Castelletti, A. (2017). "Balancing exploration, uncertainty and computational demands in many objective reservoir optimization." *Advances in Water Resources*, 109, 196–210.

Zhang, J., Wang, X., Liu, P., et al. (2017). "Assessing the weighted multi-objective adaptive surrogate model optimization to derive large-scale reservoir operating rules with sensitivity analysis." *Journal of Hydrology*, 544, 613–627.

Zhang, Q. and Li, H. (2007). "Moea/d: A multiobjective evolutionary algo-

rithm based on decomposition." *Evolutionary Computation, IEEE Transactions on*, 11(6), 712–731.

Zitzler, E. and Künzli, S. (2004). "Indicator-based selection in multiobjective search." *Parallel Problem Solving from Nature-PPSN VIII*, Springer, 832–842.

Zitzler, E., Laumanns, M., and Thiele, L. (2001). "Spea2: Improved the performance of the strength pareto evolutionary algorithm." *Report no.*, Technical Report 103, Computer Engineering and Communication Networks Lac (TIK), Swiss Federal institute of Technology (ETH) Zurich.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and Da Fonseca, V. G. (2003a). "Performance assessment of multiobjective optimizers: An analysis and review." *IEEE Transactions on evolutionary computation*, 7(2), 117–132.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and Fonseca, V. G. d. (2003b). "Performance assessment of multiobjective optimizers: an analysis and review." *IEEE Transactions on Evolutionary Computation*, 7(2), 117–132.